Неявная кросс-валидация для выбора подмножества информативных признаков в задаче обучения распознаванию образов по методу опорных векторов

Левдик Павел Владимирович
Московский физико-технический институт
Факультет управления и прикладной математики
Интеллектуальные системы

Научный руководитель
Моттль Вадим Вячеславович
Вычислительный центр РАН

Типовая задача восстановления зависимости в множестве объектов реального мира

Некоторое множество реально существующих объектов $\omega \in \Omega$ с их скрытыми характеристиками $y(\omega): \Omega \to \mathbb{Y}$.

Типовая задача восстановления зависимости в множестве объектов реального мира

Некоторое множество реально существующих объектов $\omega \in \Omega$ с их скрытыми характеристиками $y(\omega): \Omega \to \mathbb{Y}$.

Доступна только конечная обучающая совокупность $\{(\omega_j, y_j), \ j = 1, ..., N\}$

Требуется найти решающее правило оценивания скрытой характеристики для других объектов $\hat{y}(\omega): \Omega \to \mathbb{Y}$, и не слишком часто ошибаться $\hat{y}(\omega) \neq y(\omega)$.

Типовая задача восстановления зависимости в множестве объектов реального мира

Некоторое множество реально существующих объектов $\omega \in \Omega$ с их скрытыми характеристиками $y(\omega): \Omega \to \mathbb{Y}$.

Доступна только конечная обучающая совокупность $\left\{(\omega_j, y_j),\ j = 1,...,N\right\}$

Требуется найти решающее правило оценивания скрытой характеристики для других объектов $\hat{y}(\omega): \Omega \to \mathbb{Y}$, и не слишком часто ошибаться $\hat{y}(\omega) \neq y(\omega)$.

Распознавание образов $\mathbb{Y} = \{y_1,..., y_m\}$, в частности, $\mathbb{Y} = \{-1, 1\}$

Мы будем говорить о задаче распознавания образов с двумя классами объектов

Типовая задача восстановления зависимости в множестве объектов реального мира

Некоторое множество реально существующих объектов $\omega \in \Omega$ с их скрытыми характеристиками $y(\omega): \Omega \to \mathbb{Y}$.

Доступна только конечная обучающая совокупность $\left\{(\omega_j, y_j),\ j = 1, ..., N\right\}$

Требуется найти решающее правило оценивания скрытой характеристики для других объектов $\hat{y}(\omega): \Omega \to \mathbb{Y}$, и не слишком часто ошибаться $\hat{y}(\omega) \neq y(\omega)$.

Распознавание образов $\mathbb{Y} = \{y_1, ..., y_m\}$, в частности, $\mathbb{Y} = \{-1, 1\}$

Мы будем говорить о задаче распознавания образов с двумя классами объектов

Типичная для практики ситуация:

Объекты доступны для наблюдения только через их попарное сравнение

$S(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$, $\qquad S(\omega', \omega'')$ – произвольная функция парного сравнения

Типовая задача восстановления зависимости в множестве объектов реального мира

Некоторое множество реально существующих объектов $\omega \in \Omega$ с их скрытыми характеристиками $y(\omega): \Omega \to \mathbb{Y}$.

Доступна только конечная обучающая совокупность $\{(\omega_j, y_j),\ j = 1, \ldots, N\}$

Требуется найти решающее правило оценивания скрытой характеристики для других объектов $\hat{y}(\omega): \Omega \to \mathbb{Y}$, и не слишком часто ошибаться $\hat{y}(\omega) \neq y(\omega)$.

Распознавание образов $\mathbb{Y} = \{y_1, \ldots, y_m\}$, в частности, $\mathbb{Y} = \{-1, 1\}$

Мы будем говорить о задаче распознавания образов с двумя классами объектов

Типичная для практики ситуация:
Объекты доступны для наблюдения только через их попарное сравнение
$S(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$, $\qquad S(\omega', \omega'')$ – произвольная функция парного сравнения

Задачу обучения при таком предположении принято называть задачей беспризнакового распознавания образов (Featureless Pattern Recognition)

Классические подходы

- Алгоритмы вычисления оценок (Ю.И. Журавлев, 1970):
  Принцип оценивания близости объектов $B(\omega', \omega'')$.

Классические подходы

- Алгоритмы вычисления оценок (Ю.И. Журавлев, 1970):
  Принцип оценивания близости объектов $B(\omega', \omega'')$.

- Реляционный подход –Relational Discriminant Analysis (Robert Duin, 1999):
  Вектор вторичных признаков объекта $\omega \in \Omega$ относительно обучающей совокупности $\boldsymbol{x}(\omega) = \left(x_i(\omega) = S(\omega_i, \omega), i = 1, ..., N \right)$ с последующим поиском разделяющей
  гиперплоскости в $\mathbb{R}^N$ $\quad \boldsymbol{a}^T \boldsymbol{x} + b = \sum_{i=1}^{N} a_i x_i + b \begin{cases} > 0 \Rightarrow \hat{y} = 1, \\ < 0 \Rightarrow \hat{y} = -1. \end{cases}$
  Число вторичных признаков огромно.

Классические подходы

- Алгоритмы вычисления оценок (Ю.И. Журавлев, 1970):
  Принцип оценивания близости объектов $B(\omega', \omega'')$.

- Реляционный подход –Relational Discriminant Analysis (Robert Duin, 1999):
  Вектор вторичных признаков объекта $\omega \in \Omega$ относительно обучающей совокупности $\boldsymbol{x}(\omega) = \left(x_i(\omega) = S(\omega_i, \omega), i = 1, \ldots, N \right)$ с последующим поиском разделяющей гиперплоскости в $\mathbb{R}^N$ $\quad \boldsymbol{a}^T \boldsymbol{x} + b = \sum_{i=1}^{N} a_i x_i + b \begin{cases} > 0 \Rightarrow \hat{y} = 1, \\ < 0 \Rightarrow \hat{y} = -1. \end{cases}$
  Число вторичных признаков огромно.

- Relevance Vector Machine (Tipping & Bishop, 2000):
  Отбирает наиболее информативные (релевантные) объекты обучающей совокупности для сравнения, но критерий обучения не является выпуклым.

- Мы модифицировали Doubly Regularized SVM (L. Wang, J. Zhu, H. Zou):
  Выпуклый критерий обучения содержит структурный параметр селективности, управляющий числом активных вторичных признаков.

Классические подходы

- Алгоритмы вычисления оценок (Ю.И. Журавлев, 1970):
  Принцип оценивания близости объектов $B(\omega', \omega'')$.

- Реляционный подход –Relational Discriminant Analysis (Robert Duin, 1999):
  Вектор вторичных признаков объекта $\omega \in \Omega$ относительно обучающей совокупности $\boldsymbol{x}(\omega) = \big(x_i(\omega) = S(\omega_i, \omega), i = 1, \ldots, N \big)$ с последующим поиском разделяющей гиперплоскости в $\mathbb{R}^N$ $\quad \boldsymbol{a}^T \boldsymbol{x} + b = \sum_{i=1}^{N} a_i x_i + b \begin{cases} > 0 \Rightarrow \hat{y} = 1, \\ < 0 \Rightarrow \hat{y} = -1. \end{cases}$
  Число вторичных признаков огромно.

- Relevance Vector Machine (Tipping & Bishop, 2000):
  Отбирает наиболее информативные (релевантные) объекты обучающей совокупности для сравнения, но критерий обучения не является выпуклым.

- Мы модифицировали Doubly Regularized SVM (L. Wang, J. Zhu, H. Zou):
  Выпуклый критерий обучения содержит структурный параметр селективности, управляющий числом активных вторичных признаков.

Проблема: Как выбрать параметр селективности по единственной обучающей совокупности без прямого перебора в ходе кросс-валидации?

Классические подходы

- Алгоритмы вычисления оценок (Ю.И. Журавлев, 1970):
Принцип оценивания близости объектов $B(\omega', \omega'')$.

- Реляционный подход –Relational Discriminant Analysis (Robert Duin, 1999):
Вектор вторичных признаков объекта $\omega \in \Omega$ относительно обучающей совокупности $\boldsymbol{x}(\omega) = \left(x_i(\omega) = S(\omega_i, \omega), i = 1, ..., N \right)$ с последующим поиском разделяющей гиперплоскости в $\mathbb{R}^N$ $\quad \boldsymbol{a}^T \boldsymbol{x} + b = \sum_{i=1}^{N} a_i x_i + b \begin{cases} > 0 \Rightarrow \hat{y} = 1, \\ < 0 \Rightarrow \hat{y} = -1. \end{cases}$
Число вторичных признаков огромно.

- Relevance Vector Machine (Tipping & Bishop, 2000):
Отбирает наиболее информативные (релевантные) объекты обучающей совокупности для сравнения, но критерий обучения не является выпуклым.

- Мы модифицировали Doubly Regularized SVM (L. Wang, J. Zhu, H. Zou):
Выпуклый критерий обучения содержит структурный параметр селективности, управляющий числом активных вторичных признаков.

Проблема: Как выбрать параметр селективности по единственной обучающей совокупности без прямого перебора в ходе кросс-валидации?

Наше решение: Использовать принцип неявной кросс-валидации, разработанный Леной Черноусовой, для задач линейной регрессии.

Критерий обучения (квадратичная версия SVM)

Критерий обучения:

Вторичные признаки $x_{ij}(\omega)=S(\omega_i,\omega_j),\ i\in\mathbb{I}=\{1,...,n\},\ n=N,$

$$\begin{pmatrix}\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\\ \hat{b}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\end{pmatrix}:\begin{cases}\sum_{i\in\hat{\mathbb{I}}}a_i^2+C\sum_{j\in\mathbb{J}}\xi_j^2\to\min(a_i,i\in\hat{\mathbb{I}},b,\xi_j,\ j\in\mathbb{J}),\\ y_j\left(\sum_{i\in\hat{\mathbb{I}}}a_i x_{ij}+b\right)\geq 1-\xi_j,\ \xi_j\geq 0,\ j\in\mathbb{J}=\{1,...,N\},\end{cases}$$

Структурные параметры:

1) подмножество признаков (в нашем случае, вторичных)

$\hat{\mathbb{I}}\subseteq\mathbb{I}=\{1,...,n\},\ \hat{n}=|\hat{\mathbb{I}}|\leq n,\ n=N,\ \boldsymbol{a},\boldsymbol{x}_j\in\mathbb{R}^{\hat{n}}$

2) коэффициент баланса C

Критерий обучения (квадратичная версия SVM)

Критерий обучения:

Вторичные признаки $x_{ij}(\omega)=S(\omega_i,\omega_j),\ i\in\mathbb{I}=\{1,\dots,n\},\ n=N,$

$$\begin{pmatrix}\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\\ \hat{b}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\end{pmatrix}:\begin{cases}\sum_{i\in\hat{\mathbb{I}}}a_i^2+C\sum_{j\in\mathbb{J}}\xi_j^2\to\min(a_i,i\in\hat{\mathbb{I}},b,\xi_j,\ j\in\mathbb{J}),\\ \underbrace{y_j\left(\sum_{i\in\hat{\mathbb{I}}}a_ix_{ij}+b\right)\geq 1-\xi_j},\ \xi_j\geq 0,\ j\in\mathbb{J}=\{1,\dots,N\},\end{cases}$$

Структурные параметры:

1) подмножество признаков (в нашем случае, вторичных)

$\hat{\mathbb{I}}\subseteq\mathbb{I}=\{1,\dots,n\},\ \hat{n}=|\hat{\mathbb{I}}|\leq n,\ n=N,\ \boldsymbol{a},\boldsymbol{x}_j\in\mathbb{R}^{\hat{n}}$

2) коэффициент баланса C

Решение двойственной задачи:

множители Лагранжа $\underbrace{\hat{\xi}_{1,\hat{\mathbb{I}},C},\dots,\hat{\xi}_{N,\hat{\mathbb{I}},C}}$.

$$\begin{cases}\sum_{j=1}^{N}\xi_j-\frac{C}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}y_jy_l\left(\sum_{i\in\hat{\mathbb{I}}}x_{ij}x_{il}\right)\xi_j\xi_l-\frac{1}{2}\sum_{j=1}^{N}\xi_j^2\to\max_{\xi_1,\dots,\xi_N},\\ \sum_{j=1}^{N}y_j\xi_j=0,\quad \xi_j\geq 0,\ j=1,\dots,N.\end{cases}$$

Задача квадратичного программирования

Критерий обучения (квадратичная версия SVM)

Критерий обучения:

Вторичные признаки $x_{ij}(\omega)=S(\omega_i,\omega_j),\ i\in\mathbb{I}=\{1,\dots,n\},\ n=N,$

$$\begin{pmatrix}\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\\ \hat{b}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\end{pmatrix}:\begin{cases}\sum_{i\in\hat{\mathbb{I}}}a_i^2+C\sum_{j\in\mathbb{J}}\xi_j^2\to\min(a_i,i\in\hat{\mathbb{I}},b,\xi_j,\ j\in\mathbb{J}),\\ y_j\left(\sum_{i\in\hat{\mathbb{I}}}a_ix_{ij}+b\right)\ge1-\xi_j,\ \xi_j\ge0,\ j\in\mathbb{J}=\{1,\dots,N\},\end{cases}$$

Структурные параметры:

1) подмножество признаков (в нашем случае, вторичных)

$\hat{\mathbb{I}}\subseteq\mathbb{I}=\{1,\dots,n\},\ \hat{n}=|\hat{\mathbb{I}}|\le n,\ n=N,\ \boldsymbol{a},\boldsymbol{x}_j\in\mathbb{R}^{\hat{n}}$

2) коэффициент баланса C

Решение двойственной задачи:

множители Лагранжа $\hat{\xi}_{1,\hat{\mathbb{I}},C},\dots,\hat{\xi}_{N,\hat{\mathbb{I}},C}$.

Подмножество опорных объектов:

$\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}=\left\{j\in\mathbb{J}:\hat{\xi}_{\hat{\mathbb{I}},C,j}>0\right\}$ объекты с ненулевой ошибкой

$$\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}=C\sum_{j=1}^{N}y_j\hat{\xi}_{\hat{\mathbb{I}},C,j}\boldsymbol{x}_j$$

$$\begin{cases}\sum_{j=1}^{N}\xi_j-\frac{C}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}y_jy_l\left(\sum_{i\in\hat{\mathbb{I}}}x_{ij}x_{il}\right)\xi_j\xi_l-\frac{1}{2}\sum_{j=1}^{N}\xi_j^2\to\max_{\xi_1,\dots,\xi_N},\\ \sum_{j=1}^{N}y_j\xi_j=0,\quad \xi_j\ge0,\ j=1,\dots,N.\end{cases}$$

Задача квадратичного программирования

Критерий обучения (квадратичная версия SVM)

Критерий обучения:

Вторичные признаки $x_{ij}(\omega)=S(\omega_i,\omega_j)$, $i\in\mathbb{I}=\{1,...,n\}$, $n=N$,

$$\begin{pmatrix}\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\\ \hat{b}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\end{pmatrix}:\begin{cases}\sum_{i\in\hat{\mathbb{I}}}a_i^2+C\sum_{j\in\mathbb{J}}\xi_j^2\to\min(a_i,i\in\hat{\mathbb{I}},b,\xi_j,\,j\in\mathbb{J}),\\ \underbrace{y_j\left(\sum_{i\in\hat{\mathbb{I}}}a_ix_{ij}+b\right)\ge 1-\xi_j,}\ \xi_j\ge 0,\ j\in\mathbb{J}=\{1,...,N\},\end{cases}$$

Структурные параметры:

1) подмножество признаков (в нашем случае, вторичных)

$\hat{\mathbb{I}}\subseteq\mathbb{I}=\{1,...,n\}$, $\hat{n}=|\hat{\mathbb{I}}|\le n$, $n=N$, $\boldsymbol{a},\boldsymbol{x}_j\in\mathbb{R}^{\hat{n}}$

2) коэффициент баланса C

Критерий неявной кросс-валидации для оценивания обобщающей способности опирается на число опорных объектов в результате обучения $\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}$.

Решение двойственной задачи:

множители Лагранжа $\underbrace{\hat{\xi}_{1,\hat{\mathbb{I}},C},...,\hat{\xi}_{N,\hat{\mathbb{I}},C}}$.

$$\begin{cases}\sum_{j=1}^N\xi_j-\frac{C}{2}\sum_{j=1}^N\sum_{l=1}^N y_jy_l\left(\sum_{i\in\hat{\mathbb{I}}}x_{ij}x_{il}\right)\xi_j\xi_l-\frac{1}{2}\sum_{j=1}^N\xi_j^2\to\max_{\xi_1,...,\xi_N},\\ \sum_{j=1}^N y_j\xi_j=0,\quad \xi_j\ge 0,\ j=1,...,N.\end{cases}$$

Задача квадратичного программирования

Подмножество опорных объектов:

$\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}=\left\{j\in\mathbb{J}:\hat{\xi}_{\hat{\mathbb{I}},C,j}>0\right\}$ объекты с ненулевой ошибкой

$$\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}=C\sum_{j=1}^N y_j\hat{\xi}_{\hat{\mathbb{I}},C,j}\boldsymbol{x}_j$$

Критерий обучения (квадратичная версия SVM)

Критерий обучения:

Вторичные признаки $x_{ij}(\omega)=S(\omega_i,\omega_j),\ i\in\mathbb{I}=\{1,...,n\},\ n=N,$

$$\begin{pmatrix}\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\\ \hat{b}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\end{pmatrix}:\begin{cases}\sum_{i\in\hat{\mathbb{I}}}a_i^2+C\sum_{j\in\mathbb{J}}\xi_j^2\to\min(a_i,i\in\hat{\mathbb{I}},b,\xi_j,\ j\in\mathbb{J}),\\ \underbrace{y_j\left(\sum_{i\in\hat{\mathbb{I}}}a_ix_{ij}+b\right)\ge1-\xi_j,}\ \xi_j\ge0,\ j\in\mathbb{J}=\{1,...,N\},\end{cases}$$

Структурные параметры:

1) подмножество признаков (в нашем случае, вторичных)

$\hat{\mathbb{I}}\subseteq\mathbb{I}=\{1,...,n\},\ \hat{n}=|\hat{\mathbb{I}}|\le n,\ n=N,\ \boldsymbol{a},\boldsymbol{x}_j\in\mathbb{R}^{\hat{n}}$

2) коэффициент баланса C

Критерий неявной кросс-валидации для оценивания обобщающей способности опирается на число опорных объектов в результате обучения $\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}$.

Решение двойственной задачи:

множители Лагранжа $\overbrace{\hat{\xi}_{1,\hat{\mathbb{I}},C},...,\hat{\xi}_{N,\hat{\mathbb{I}},C}}$.

$$\begin{cases}\sum_{j=1}^{N}\xi_j-\frac{C}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}y_jy_l\left(\sum_{i\in\hat{\mathbb{I}}}x_{ij}x_{il}\right)\xi_j\xi_l-\frac{1}{2}\sum_{j=1}^{N}\xi_j^2\to\max_{\xi_1,...,\xi_N},\\ \sum_{j=1}^{N}y_j\xi_j=0,\quad \xi_j\ge0,\ j=1,...,N.\end{cases}$$

Подмножество опорных объектов:

$\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}=\left\{j\in\mathbb{J}:\hat{\xi}_{\hat{\mathbb{I}},C,j}>0\right\}$ объекты с ненулевой ошибкой

$\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}=C\sum_{j=1}^{N}y_j\hat{\xi}_{\hat{\mathbb{I}},C,j}\boldsymbol{x}_j$

Задача квадратичного программирования

Предложенный в диссертации критерий неявной кросс-валидации для выбора подмножества признаков с наилучшей обобщающей способностью обучения

$$\begin{pmatrix}\hat{\mathbb{I}}(\boldsymbol{y},\boldsymbol{X})\\ \hat{C}(\boldsymbol{y},\boldsymbol{X})\end{pmatrix}=\arg\min_{\hat{\mathbb{I}}\subseteq\mathbb{I},C}\Big\{\underbrace{\sum_{j\in\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}}\hat{\xi}^2_{\hat{\mathbb{I}},C,j}}_{\text{эмпирический риск}}+\underbrace{\Delta(\hat{\mathbb{I}},C,\boldsymbol{X})}_{\text{штраф за сложность модели}}\Big\},\ \Delta(\hat{\mathbb{I}},C,\boldsymbol{X})=Tr\left[\hat{\mathbf{X}}_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}\hat{\mathbf{X}}^T_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}\left(\hat{\mathbf{X}}_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}\hat{\mathbf{X}}^T_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}+\mathbf{B}_{\hat{\mathbb{I}},C}\right)^{-1}\right]$$

Критерий обучения (квадратичная версия SVM)

Критерий обучения:

Вторичные признаки $x_{ij}(\omega)=S(\omega_i,\omega_j)$, $i\in\mathbb{I}=\{1,\dots,n\}$, $n=N$,

$$\begin{pmatrix}\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\\ \hat{b}_{\hat{\mathbb{I}},C}(\boldsymbol{X},\boldsymbol{y})\end{pmatrix}:\begin{cases}\sum_{i\in\hat{\mathbb{I}}}a_i^2+C\sum_{j\in\mathbb{J}}\xi_j^2\to\min(a_i,i\in\hat{\mathbb{I}},b,\xi_j,\ j\in\mathbb{J}),\\ y_j\left(\sum_{i\in\hat{\mathbb{I}}}a_ix_{ij}+b\right)\ge 1-\xi_j,\ \xi_j\ge 0,\ j\in\mathbb{J}=\{1,\dots,N\},\end{cases}$$

Структурные параметры:

1) подмножество признаков (в нашем случае, вторичных)

$\hat{\mathbb{I}}\subseteq\mathbb{I}=\{1,\dots,n\}$, $\hat{n}=|\hat{\mathbb{I}}|\le n$, $n=N$, $\boldsymbol{a},\boldsymbol{x}_j\in\mathbb{R}^{\hat{n}}$

2) коэффициент баланса C

Критерий неявной кросс-валидации для оценивания обобщающей способности опирается на число опорных объектов в результате обучения $\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}$.

Решение двойственной задачи:

множители Лагранжа $\overbrace{\hat{\xi}_{1,\hat{\mathbb{I}},C},\dots,\hat{\xi}_{N,\hat{\mathbb{I}},C}}$.

$$\begin{cases}\sum_{j=1}^{N}\xi_j-\frac{C}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}y_jy_l\left(\sum_{i\in\hat{\mathbb{I}}}x_{ij}x_{il}\right)\xi_j\xi_l-\frac{1}{2}\sum_{j=1}^{N}\xi_j^2\to\max_{\xi_1,\dots,\xi_N},\\ \sum_{j=1}^{N}y_j\xi_j=0,\quad \xi_j\ge 0,\ j=1,\dots,N.\end{cases}$$

Задача квадратичного программирования

Подмножество опорных объектов:

$\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}=\left\{j\in\mathbb{J}:\hat{\xi}_{\hat{\mathbb{I}},C,j}>0\right\}$ объекты с ненулевой ошибкой

$$\hat{\boldsymbol{a}}_{\hat{\mathbb{I}},C}=C\sum_{j=1}^{N}y_j\hat{\xi}_{\hat{\mathbb{I}},C,j}\boldsymbol{x}_j$$

Предложенный в диссертации критерий неявной кросс-валидации для выбора подмножества признаков с наилучшей обобщающей способностью обучения

$$\begin{pmatrix}\hat{\mathbb{I}}(\boldsymbol{y},\boldsymbol{X})\\ \hat{C}(\boldsymbol{y},\boldsymbol{X})\end{pmatrix}=\arg\min_{\hat{\mathbb{I}}\subseteq\mathbb{I},\,C}\Big\{\underbrace{\sum_{j\in\hat{\mathbb{J}}_{\hat{\mathbb{I}},C}}\hat{\xi}^2_{\hat{\mathbb{I}},C,j}}_{\textit{эмпирический риск}}+\underbrace{\Delta(\hat{\mathbb{I}},C,\boldsymbol{X})}_{\textit{штраф за сложность модели}}\Big\},\quad \Delta(\hat{\mathbb{I}},C,\boldsymbol{X})=Tr\left[\hat{\mathbf{X}}_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}\hat{\mathbf{X}}^T_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}\left(\hat{\mathbf{X}}_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}\hat{\mathbf{X}}^T_{\hat{\mathbb{J}}_{\hat{\mathbb{I}},C},\hat{\mathbb{I}},C}+\mathbf{B}_{\hat{\mathbb{I}},C}\right)^{-1}\right]$$

Огромное число 2^n всех подмножеств в множестве признаков $\hat{\mathbb{I}}\subseteq\mathbb{I}=\{1,\dots,n\}$ делает невозможной «наивную» реализацию этого критерия.

Покажем, как можно обойти это препятствие и обойтись вообще без перебора.

SVM с автоматическим отбором признаков

SVM с квадратичной регуляризацией:

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \ge 1-\xi_j,\ \xi_j \ge 0,\ j\in\mathbb{J}=\{1,\ldots,N\}. \end{cases}$$

SVM с автоматическим отбором признаков

SVM с квадратичной регуляризацией:

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \ge 1-\xi_j,\ \xi_j \ge 0,\ j\in\mathbb{J}=\{1,\dots,N\}. \end{cases}$$

SVM с квадратично-модульной регуляризацией

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + \boxed{\mu\sum_{i\in\mathbb{I}} |a_i|} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \ge 1-\xi_j,\ \xi_j \ge 0,\ j\in\mathbb{J}=\{1,\dots,N\}. \end{cases}$$

Параметр селективности $\mu \ge 0$

SVM с автоматическим отбором признаков

SVM с квадратичной регуляризацией:

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \geq 1-\xi_j,\ \xi_j \geq 0,\ j\in\mathbb{J}=\{1,\ldots,N\}. \end{cases}$$

SVM с квадратично-модульной регуляризацией

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + \boxed{\mu\sum_{i\in\mathbb{I}} |a_i|} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \geq 1-\xi_j,\ \xi_j \geq 0,\ j\in\mathbb{J}=\{1,\ldots,N\}. \end{cases}$$

Параметр селективности $\mu \geq 0$

Двойственная задача:

$$\begin{cases} \sum_{j=1}^{N} \xi_j - \frac{C}{2}\sum_{i\in\mathbb{I}}\left\{\min\left[\mu + \sum^{N} y_j\xi_j x_{ij},\ 0,\ \mu - \sum_{j=1}^{N} y_j\xi_j x_{ij}\right]\right\}^2 - 2C^2\sum_{j=1}^{N}\xi_j^2 \to \max(\xi_1,\ldots,\xi_N), \\ \sum_{j=1}^{N} y_j\xi_j = 0, \quad \xi_j \geq 0, \quad j = 1,\ldots,N. \end{cases}$$

Задача выпуклого программирования
Алгоритм внутренней точки

Решение двойственной задачи: множители Лагранжа $\hat{\xi}_{C,\mu,1},\ldots,\hat{\xi}_{C,\mu,N}$.

SVM с автоматическим отбором признаков

SVM с квадратичной регуляризацией:

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \geq 1-\xi_j,\ \xi_j \geq 0,\ j\in\mathbb{J}=\{1,\dots,N\}. \end{cases}$$

SVM с квадратично-модульной регуляризацией

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + \boxed{\mu\sum_{i\in\mathbb{I}} |a_i|} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \geq 1-\xi_j,\ \xi_j \geq 0,\ j\in\mathbb{J}=\{1,\dots,N\}. \end{cases}$$

Параметр селективности $\mu \geq 0$

Двойственная задача:

$$\begin{cases} \sum_{j=1}^{N} \xi_j - \frac{C}{2}\sum_{i\in\mathbb{I}}\left\{\min\left[\mu+\sum^{N} y_j\xi_j x_{ij},\ 0,\ \mu-\sum_{j=1}^{N} y_j\xi_j x_{ij}\right]\right\}^2 - 2C^2\sum_{j=1}^{N}\xi_j^2 \to \max(\xi_1,\dots,\xi_N), \\ \sum_{j=1}^{N} y_j\xi_j = 0, \quad \xi_j \geq 0, \quad j=1,\dots,N. \end{cases}$$

Задача выпуклого программирования
Алгоритм внутренней точки

Решение двойственной задачи: множители Лагранжа $\hat{\xi}_{C,\mu,1},\dots,\hat{\xi}_{C,\mu,N}$.

$$\hat{a}_{C,\mu,i} = \begin{cases} \left(\sum_{j\in\hat{\mathbb{J}}_{\beta,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij}+\mu\right)<0, & \sum_{j\in\hat{\mathbb{J}}_{C,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij} < -\frac{\mu}{2}, \\ 0, & -\frac{\mu}{2} \leq \sum_{j\in\hat{\mathbb{J}}_{C,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij} \leq \frac{\mu}{2}, \\ \left(\sum_{j\in\hat{\mathbb{J}}_{\beta,\mu}} y_j\hat{\xi}_j x_{ij}-\mu\right)>0, & \sum_{j\in\hat{\mathbb{J}}_{C,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij} > \frac{\mu}{2}. \end{cases}$$

SVM с автоматическим отбором признаков

SVM с квадратичной регуляризацией:

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \geq 1-\xi_j,\ \xi_j \geq 0,\ j\in\mathbb{J}=\{1,\dots,N\}. \end{cases}$$

SVM с квадратично-модульной регуляризацией

$$\begin{cases} \boxed{\sum_{i\in\mathbb{I}} a_i^2} + \boxed{\mu\sum_{i\in\mathbb{I}} |a_i|} + C\sum_{j\in\mathbb{J}} \xi_j^2 \to \min(a_i, i\in\mathbb{I}, b, \xi_j, j\in\mathbb{J}), \\ y_j\left(\sum_{i\in\mathbb{I}} a_i x_{ij} + b\right) \geq 1-\xi_j,\ \xi_j \geq 0,\ j\in\mathbb{J}=\{1,\dots,N\}. \end{cases}$$

Параметр селективности $\mu \geq 0$

Двойственная задача:

$$\begin{cases} \sum_{j=1}^N \xi_j - \frac{C}{2}\sum_{i\in\mathbb{I}}\left\{\min\left[\mu + \sum^N y_j\xi_j x_{ij},\ 0,\ \mu - \sum_{j=1}^N y_j\xi_j x_{ij}\right]\right\}^2 - 2C^2\sum_{j=1}^N \xi_j^2 \to \max(\xi_1,\dots,\xi_N), \\ \sum_{j=1}^N y_j\xi_j = 0, \quad \xi_j \geq 0, \quad j=1,\dots,N. \end{cases}$$

Задача выпуклого программирования
Алгоритм внутренней точки

Решение двойственной задачи: множители Лагранжа $\hat{\xi}_{C,\mu,1},\dots,\hat{\xi}_{C,\mu,N}$.

Другой аспект решения – отбор признаков:
чем больше μ, тем меньше признаков остаются активными.

$$\hat{a}_{C,\mu,i} = \begin{cases} \left(\sum_{j\in\hat{\mathbb{J}}_{\beta,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij} + \mu\right) < 0, & \sum_{j\in\hat{\mathbb{J}}_{C,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij} < -\frac{\mu}{2}, \\ 0, & -\frac{\mu}{2} \leq \sum_{j\in\hat{\mathbb{J}}_{C,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij} \leq \frac{\mu}{2}, \\ \left(\sum_{j\in\hat{\mathbb{J}}_{\beta,\mu}} y_j\hat{\xi}_j x_{ij} - \mu\right) > 0, & \sum_{j\in\hat{\mathbb{J}}_{C,\mu}} y_j\hat{\xi}_{C,\mu,j}x_{ij} > \frac{\mu}{2}. \end{cases}$$

Подмножество опорных признаков:

$$\hat{\mathbb{I}}_{C,\mu} = \left\{i\in\mathbb{I}: \hat{a}_{C,\mu,i} \neq 0\right\} \subseteq \hat{\mathbb{I}} = \{1,\dots,n\},\ n = N.$$

$$\mu = 0 \Rightarrow \hat{\mathbb{I}}_{C,\mu} = \hat{\mathbb{I}} = \{1,\dots,N\},$$

$$\mu \to \infty \Rightarrow \hat{\mathbb{I}}_{C,\mu} \to \varnothing.$$

Алгоритм выбора подмножества вторичных признаков

Обучающая совокупность $\left\{(\omega_j, y_j = \pm 1),\ j = 1, \dots, N\right\}$. Полное множество объектов $\mathbb{J} = \{1, \dots, N\}$.

Полное множество вторичных признаков $\mathbb{I} = \{1, \dots, n\}$, $n = N$, $\quad \boldsymbol{x}_j = \left(x_{ij}(\omega) = S(\omega_i, \omega_j),\ i \in \mathbb{I}\right)$.

Алгоритм выбора подмножества вторичных признаков

Обучающая совокупность $\left\{(\omega_j, y_j = \pm 1),\ j = 1, \dots, N\right\}$. Полное множество объектов $\mathbb{J} = \{1, \dots, N\}$.
Полное множество вторичных признаков $\mathbb{I} = \{1, \dots, n\}$, $n = N$, $\boldsymbol{x}_j = \left(x_{ij}(\omega) = S(\omega_i, \omega_j),\ i \in \mathbb{I}\right)$.
1) Последовательность возрастающих значений параметра селективности $\mu_0 = 0 < \mu_1 < \dots < \mu_m$.

Алгоритм выбора подмножества вторичных признаков

Обучающая совокупность $\left\{(\omega_j, y_j = \pm 1),\ j = 1, ..., N\right\}$. Полное множество объектов $\mathbb{J} = \{1, ..., N\}$.

Полное множество вторичных признаков $\mathbb{I} = \{1, ..., n\}$, $n = N$, $\quad \boldsymbol{x}_j = \left(x_{ij}(\omega) = S(\omega_i, \omega_j),\ i \in \mathbb{I} \right)$.

1) Последовательность возрастающих значений параметра селективности $\mu_0 = 0 < \mu_1 < ... < \mu_m$.

2) Последовательность решений задачи SVM с квадратично-модульной регуляризацией $C = const \to \infty$ и последовательность уменьшающихся подмножеств опорных признаков. Обычно $\hat{\mathbb{I}}_{\mu_0} \supset \hat{\mathbb{I}}_{\mu_1} \supset ... \supset \hat{\mathbb{I}}_{\mu_m}$, но вложенность не гарантируется.

Алгоритм выбора подмножества вторичных признаков

Обучающая совокупность $\left\{(\omega_j, y_j = \pm 1),\ j = 1,...,N\right\}$. Полное множество объектов $\mathbb{J} = \{1,...,N\}$.

Полное множество вторичных признаков $\mathbb{I} = \{1,...,n\}$, $n = N$, $\quad \boldsymbol{x}_j = \left(x_{ij}(\omega) = S(\omega_i, \omega_j),\ i \in \mathbb{I}\right)$.

1) Последовательность возрастающих значений параметра селективности $\mu_0 = 0 < \mu_1 < ... < \mu_m$.

2) Последовательность решений задачи SVM с квадратично-модульной регуляризацией $C = const \to \infty$ и последовательность уменьшающихся подмножеств опорных признаков. Обычно $\hat{\mathbb{I}}_{\mu_0} \supset \hat{\mathbb{I}}_{\mu_1} \supset ... \supset \hat{\mathbb{I}}_{\mu_m}$, но вложенность не гарантируется.

3) Последовательность решений задачи SVM с квадратичной регуляризацией для найденных подмножеств признаков $\left|\hat{\mathbb{I}}_{\mu_0}\right| = \hat{n}_{\mu_0}, \left|\hat{\mathbb{I}}_{\mu_1}\right| = \hat{n}_{\mu_1}, ..., \left|\hat{\mathbb{I}}_{\mu_m}\right| = \hat{n}_{\mu_m}$.
Результат – последовательность подмножеств опорных объектов.
Как правило, $\left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_0}}\right| = \hat{N}_{\mu_0} < \left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_1}}\right| = \hat{N}_{\mu_1} < ... < \left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_m}}\right| = \hat{N}_{\mu_m}$ в силу ухудшения разделимости обучающей совокупности.
Последовательность дискриминантных гиперплоскостей в соответствующих $\hat{n}_{\mu_1}, ..., \hat{n}_{\mu_m}$-мерных пространствах $\sum_{i=1}^{N} \sum_{i \in \hat{\mathbb{I}}_{\mu_k}} \hat{a}_{C\mu_k, i} x_i + \hat{b}_{C\mu_k} \gtrless 0$.

Алгоритм выбора подмножества вторичных признаков

Обучающая совокупность $\left\{(\omega_j, y_j{=}\pm1),\ j=1,\ldots,N\right\}$. Полное множество объектов $\mathbb{J}=\{1,\ldots,N\}$.

Полное множество вторичных признаков $\mathbb{I}=\{1,\ldots,n\}$, $n=N$, $\quad \boldsymbol{x}_j=\left(x_{ij}(\omega)=S(\omega_i,\omega_j),\ i\in\mathbb{I}\right)$.

1) Последовательность возрастающих значений параметра селективности $\mu_0=0<\mu_1<\ldots<\mu_m$.

2) Последовательность решений задачи SVM с квадратично-модульной регуляризацией $C{=}const\to\infty$ и последовательность уменьшающихся подмножеств опорных признаков. Обычно $\hat{\mathbb{I}}_{\mu_0}\supset\hat{\mathbb{I}}_{\mu_1}\supset\ldots\supset\hat{\mathbb{I}}_{\mu_m}$, но вложенность не гарантируется.

3) Последовательность решений задачи SVM с квадратичной регуляризацией для найденных подмножеств признаков $\left|\hat{\mathbb{I}}_{\mu_0}\right|=\hat{n}_{\mu_0},\left|\hat{\mathbb{I}}_{\mu_1}\right|=\hat{n}_{\mu_1},\ldots,\left|\hat{\mathbb{I}}_{\mu_m}\right|=\hat{n}_{\mu_m}$.
Результат – последовательность подмножеств опорных объектов.
Как правило, $\left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_0}}\right|=\hat{N}_{\mu_0}<\left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_1}}\right|=\hat{N}_{\mu_1}<\ldots<\left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_m}}\right|=\hat{N}_{\mu_m}$ в силу ухудшения разделимости обучающей совокупности.
Последовательность дискриминантных гиперплоскостей в соответствующих $\hat{n}_{\mu_1},\ldots,\hat{n}_{\mu_m}$-мерных пространствах $\sum_{i=1}^{N}\sum_{i\in\hat{\mathbb{I}}_{\mu_k}}\hat{a}_{C\mu_k,i}x_i+\hat{b}_{C\mu_k}\gtrless0$.

4) Применение критерия неявной кросс-валидации

$$\hat{\mathbb{I}}=\underset{\hat{\mathbb{I}}\subseteq\mathbb{I},\,C}{\arg\min}\left\{\sum\nolimits_{j\in\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_k}}}\hat{\xi}^2_{\hat{\mathbb{I}}_{\mu_k},j}+\Delta(\hat{\mathbb{I}}_{\mu_k},\boldsymbol{X})\right\},\quad \Delta(\hat{\mathbb{I}}_{\mu_k},\boldsymbol{X})=\min\left\{\hat{n}_{\mu_k},\ \hat{N}_{\mu_k}-1\right\}+1,$$

поскольку в диссертации доказано, что

$$\lim_{C\to\infty}\Delta(\hat{\mathbb{I}}_{\mu_k},C,\boldsymbol{X})=\lim_{C\to\infty}Tr\left[\hat{\mathbf{X}}_{\hat{\mathbb{I}}_{\mu_k}}\hat{\mathbf{X}}^T_{\hat{\mathbb{I}}_{\mu_k}}(\hat{\mathbf{X}}_{\hat{\mathbb{I}}_{\mu_k}}\hat{\mathbf{X}}^T_{\hat{\mathbb{I}}_{\mu_k}}+\mathbf{B}_{\hat{\mathbb{I}}_{\mu_k}})^{-1}\right]=\min\left\{\hat{n}_{\hat{\mathbb{I}}_{\mu_k}},\ \hat{N}_{\hat{\mathbb{I}}_{\mu_k}}-1\right\}+1.$$

Алгоритм выбора подмножества вторичных признаков

Обучающая совокупность $\left\{(\omega_j, y_j = \pm 1),\ j = 1,\dots,N\right\}$. Полное множество объектов $\mathbb{J} = \{1,\dots,N\}$.

Полное множество вторичных признаков $\mathbb{I} = \{1,\dots,n\}$, $n = N$, $\quad \boldsymbol{x}_j = \left(x_{ij}(\omega) = S(\omega_i, \omega_j),\ i \in \mathbb{I}\right)$.

1) Последовательность возрастающих значений параметра селективности $\mu_0 = 0 < \mu_1 < \dots < \mu_m$.

2) Последовательность решений задачи SVM с квадратично-модульной регуляризацией $C = const \to \infty$ и последовательность уменьшающихся подмножеств опорных признаков. Обычно $\hat{\mathbb{I}}_{\mu_0} \supset \hat{\mathbb{I}}_{\mu_1} \supset \dots \supset \hat{\mathbb{I}}_{\mu_m}$, но вложенность не гарантируется.

3) Последовательность решений задачи SVM с квадратичной регуляризацией для найденных подмножеств признаков $\left|\hat{\mathbb{I}}_{\mu_0}\right| = \hat{n}_{\mu_0}, \left|\hat{\mathbb{I}}_{\mu_1}\right| = \hat{n}_{\mu_1}, \dots, \left|\hat{\mathbb{I}}_{\mu_m}\right| = \hat{n}_{\mu_m}$.
Результат – последовательность подмножеств опорных объектов.
Как правило, $\left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_0}}\right| = \hat{N}_{\mu_0} < \left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_1}}\right| = \hat{N}_{\mu_1} < \dots < \left|\hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_m}}\right| = \hat{N}_{\mu_m}$ в силу ухудшения разделимости обучающей совокупности.
Последовательность дискриминантных гиперплоскостей в соответствующих $\hat{n}_{\mu_1}, \dots, \hat{n}_{\mu_m}$-мерных пространствах $\sum_{i=1}^{N} \sum_{i \in \hat{\mathbb{I}}_{\mu_k}} \hat{a}_{C\mu_k, i} x_i + \hat{b}_{C\mu_k} \gtrless 0$.

4) Применение критерия неявной кросс-валидации

$$\hat{\mathbb{I}} = \arg\min_{\hat{\mathbb{I}} \subseteq \mathbb{I},\, C} \left\{ \sum_{j \in \hat{\mathbb{J}}_{\hat{\mathbb{I}}_{\mu_k}}} \hat{\xi}^2_{\hat{\mathbb{I}}_{\mu_k}, j} + \Delta(\hat{\mathbb{I}}_{\mu_k}, \boldsymbol{X}) \right\},\quad \Delta(\hat{\mathbb{I}}_{\mu_k}, \boldsymbol{X}) = \min\left\{\hat{n}_{\mu_k},\ \hat{N}_{\mu_k} - 1\right\} + 1,$$

поскольку в диссертации доказано, что

$$\lim_{C \to \infty} \Delta(\hat{\mathbb{I}}_{\mu_k}, C, \boldsymbol{X}) = \lim_{C \to \infty} Tr\left[\hat{\mathbf{X}}_{\hat{\mathbb{I}}_{\mu_k}} \hat{\mathbf{X}}^T_{\hat{\mathbb{I}}_{\mu_k}} (\hat{\mathbf{X}}_{\hat{\mathbb{I}}_{\mu_k}} \hat{\mathbf{X}}^T_{\hat{\mathbb{I}}_{\mu_k}} + \mathbf{B}_{\hat{\mathbb{I}}_{\mu_k}})^{-1} \right] = \min\left\{\hat{n}_{\hat{\mathbb{I}}_{\mu_k}},\ \hat{N}_{\hat{\mathbb{I}}_{\mu_k}} - 1\right\} + 1.$$

Для сравнения мы применили также традиционную переборную кросс-валидацию – метод скользящего контроля.

Источник экспериментального массива данных

Chicken Pieces Silhouettes Database http://algoval.essex.ac.uk/data/sequence/chicken/

Массив содержит 446 силуэтных изображений 5-ти частей тушки курицы.

$\Omega = \{\omega_1, ..., \omega_{446}\}, \ldots\ y(\omega) \in \mathbb{Y} = \{1, 2, 3, 4, 5\}$.

Источник экспериментального массива данных

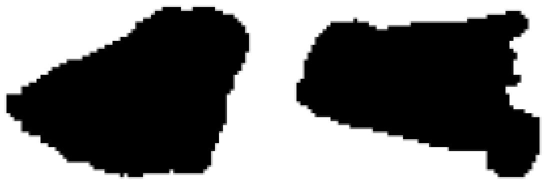
Chicken Pieces Silhouettes Database http://algoval.essex.ac.uk/data/sequence/chicken/

Массив содержит 446 силуэтных изображений 5-ти частей тушки курицы.

$\Omega = \{\omega_1, \dots, \omega_{446}\}, \dots\ y(\omega) \in \mathbb{Y} = \{1, 2, 3, 4, 5\}$.

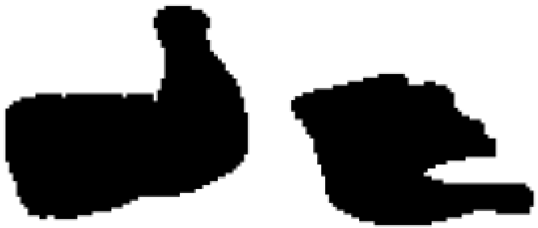
Размер и ориентация изображений безразличны.

(1) Крылышко 117



(2) Спинка 76

(3) Ножка 96



(4) Бедрышко 61

(5) Грудка 96

Источник экспериментального массива данных

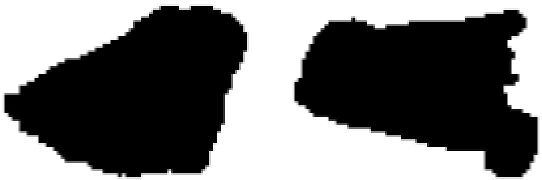
Chicken Pieces Silhouettes Database http://algoval.essex.ac.uk/data/sequence/chicken/

Массив содержит 446 силуэтных изображений 5-ти частей тушки курицы.

$\Omega = \{\omega_1, ..., \omega_{446}\}, \dots\ y(\omega) \in \mathbb{Y} = \{1, 2, 3, 4, 5\}$.

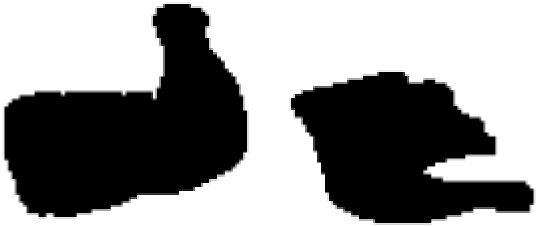
Размер и ориентация изображений безразличны.

(1) Крылышко 117



(2) Спинка 76

(3) Ножка 96



(4) Бедрышко 61

(5) Грудка 96

На множестве изображений определены 44 разные числовые функции парного сравнения изображений, представляющие их в компьютере $S_i(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$, $i = 1, ..., 44$,

Каждая функция выражает некоторую меру сходства контуров сравниваемых изображений.

Источник экспериментального массива данных

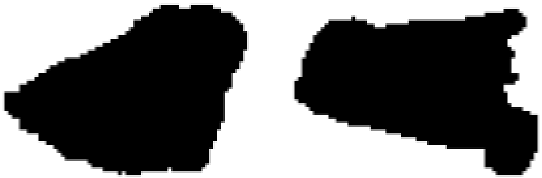
Chicken Pieces Silhouettes Database   http://algoval.essex.ac.uk/data/sequence/chicken/

Массив содержит 446 силуэтных изображений 5-ти частей тушки курицы.

$\Omega = \{\omega_1, ..., \omega_{446}\}, \ldots \, y(\omega) \in \mathbb{Y} = \{1, 2, 3, 4, 5\}$.

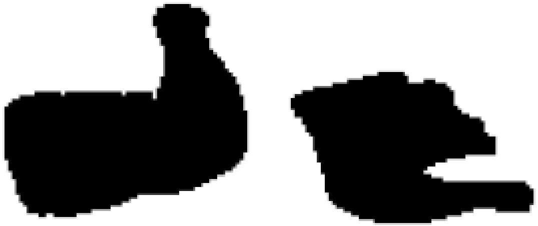
Размер и ориентация изображений безразличны.

(1) Крылышко 117



(2) Спинка 76

(3) Ножка 96



(4) Бедрышко 61

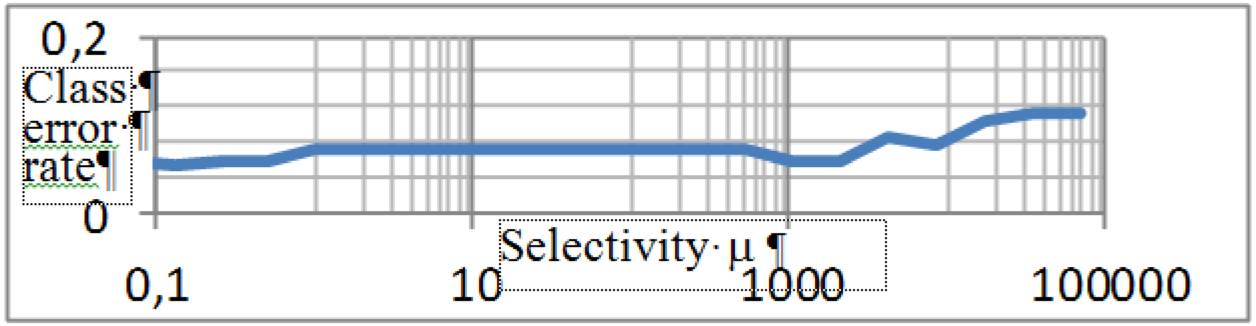
(5) Грудка 96

На множестве изображений определены 44 разные числовые функции парного сравнения изображений, представляющие их в компьютере $S_i(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$, $i = 1, ..., 44$,
Каждая функция выражает некоторую меру сходства контуров сравниваемых изображений.

В наших экспериментах мы пока использовали только два класса изображений (2)-(3)
$\Omega = \left\{\omega_j,\ j = 1, ..., N = 172\right\},\ y_j = \pm 1$, и одну функцию парного сравнения $S(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$.
Каждый объект представлен N-мерным вектором его вторичных признаков
$\mathbf{x}_j = (x_{i1},\ i \in \mathbb{I}) = \big(S(\omega_i, \omega_1),\ i \in \mathbb{I}\big) \in \mathbb{R}^{n=N}$, $\mathbb{I} = \{1, ..., n\}$ полное множество вторичных признаков.

Сравнение традиционной и неявной кросс-валидации

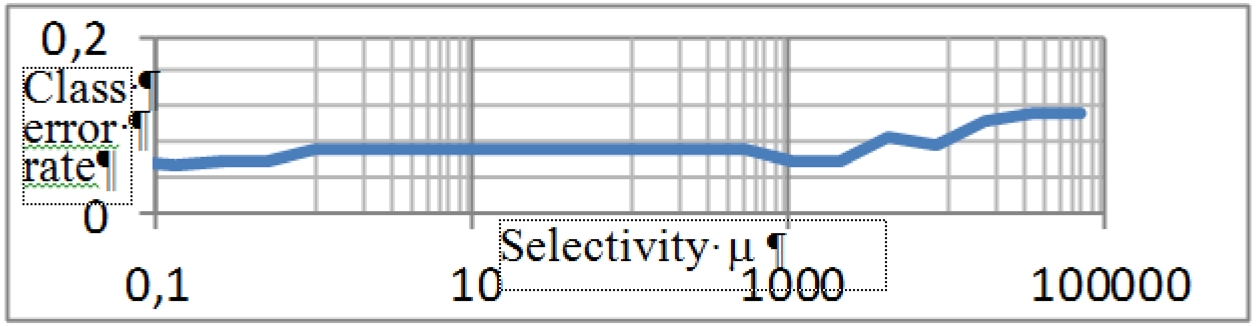
Обычный метод скользящего контроля:



Оценка обобщающей способности по методу традиционного скользящего контроля практически не зависит от уровня селективности. Это означает, что стандартный метод SVM без отбора признаков ($\mu = 0$) не склонен к переобучению, даже если использовать все огромное множество вторичных признаков.

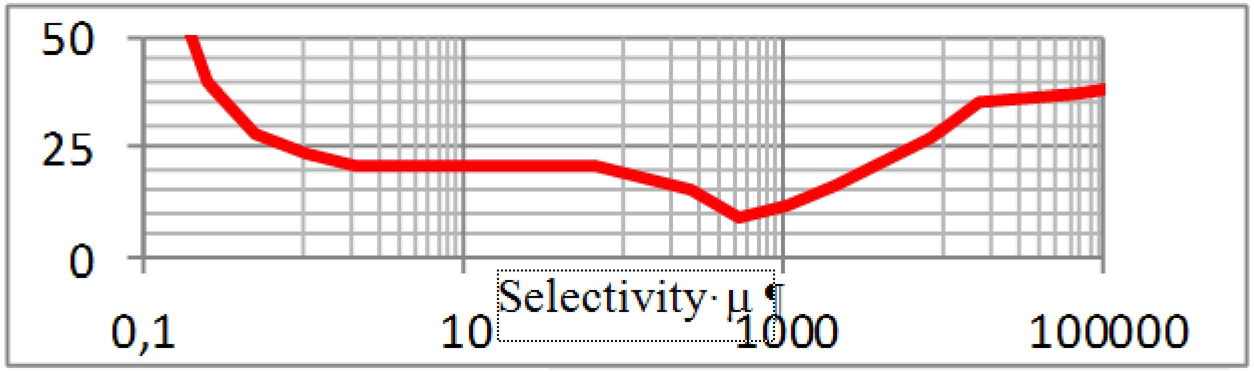
Сравнение традиционной и неявной кросс-валидации

Обычный метод скользящего контроля:



Оценка обобщающей способности по методу традиционного скользящего контроля практически не зависит от уровня селективности. Это означает, что стандартный метод SVM без отбора признаков ($\mu = 0$) не склонен к переобучению, даже если использовать все огромное множество вторичных признаков.

Неявная кросс-валидация:



Наблюдается отчетливый минимум структурного риска при $\mu = 500$. В этой точке неявная кросс-валидация выделяет подмножество из всего лишь 5-ти релевантных объектов, достаточное для решающего правила классификации новых объектов.

Заключение

1) Побудительным мотивом для создания альтернативного метода кросс-валидации было стремление избежать многократного повторения процедур обучения и контроля качества решающего правила при различных разбиениях исходной выборки.
2) Разработан метод математической имитации кросс-валидации, основанный на естественных допущениях о природе неизвестного вероятностного распределения, породившего обучающую выборку. В нашем методе многократное повторение процесса обучения заменяется вычислением единственного критерия, зависящего только от исходных данных и параметров регуляризации.
3) Эксперимент показал, что, кроме сокращения объема вычислений, метод неявной кросс-валидации обладает существенно иными свойствами по сравнению с традиционной переборной кросс-валидацией.