

Математические методы распознавания образов

Виктор Владимирович Китов

1 сентября 2016 г.

Информация о курсе

- Лектор - Виктор Владимирович Китов, семинарист - Евгений Соколов
- 12-14 лекций
- семинары (задачи, программирование на python)
- подробности на сайте machinelearning.ru, в поиске раздел «Машинное обучение (курс лекций, В.В.Китов)/2015-2016».

Рекомендуемые материалы по курсу

- **Курс лекций по машинному обучению.** Константин Воронцов. См. machinelearning.ru.
- **Statistical Pattern Recognition.** 3rd Edition, Andrew R. Webb, Keith D. Copsey, John Wiley & Sons Ltd., 2011.
- **The Elements of Statistical Learning: Data Mining, Inference, and Prediction.** Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2nd Edition, Springer, 2009.
<http://statweb.stanford.edu/~tibs/ElemStatLearn/>.
- **Machine Learning: A Probabilistic Perspective.** Kevin P. Murphy. Massachusetts Institute of Technology. 2012.
- **Pattern Recognition and Machine Learning.** Christopher M. Bishop. Springer. 2006.
- **Дополнительные источники** - wikipedia, статьи, видео-лекции (в основном, на английском).

Содержание

- 1 Задачи распознавания образов
- 2 Основные понятия задачи обучения с учителем.
- 3 Практическое применение алгоритмов машинного обучения

Распознавание образов и машинное обучение

- Теория распознавания образов — раздел информатики и смежных дисциплин, развивающий основы и методы классификации и идентификации предметов, процессов, сигналов, ситуаций и аналогичных явлений, которые характеризуются конечным набором некоторых свойств и признаков.
- Машинное обучение - наука об алгоритмах, которые сами настраиваются на данных.
- Распознавание образов использует методы машинного обучения, созданные для восстановления зависимости между признаками и распознаваемыми явлениями.
- В дальнейшем, под машинным обучением будет пониматься класс всех методов машинного обучения, используемых для распознавания образов, и два термина будут взаимозаменяемы.

Общая задача распознавания образов

- Рассматривается множество объектов O
- Каждый объект описывается набором известных признаков $x \in \mathcal{X}$ и характеристик $y \in \mathcal{Y}$.

$$o \in O \longrightarrow (x, y)$$

- Обычно $\mathcal{X} = \mathbb{R}^D$, \mathcal{Y} - число, но в общем случае - любые структурные описания объектов.

Общая задача распознавания образов

- Задача: найти отображение f , которое бы приближало $\mathcal{X} \rightarrow \mathcal{Y}$.
 - по известным данным о конечном числе объектов
 - на интересующем наборе других объектов
- Вопросы, которые решаются в машинном обучении:
 - как выбрать признаки и характеристики
 - в каком смысле отображение a должно приближать реальную взаимосвязь
 - как строить f

Варианты постановок задач

- Для каждого нового объекта x требуется сопоставить y .
- Что известно:
 - $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ - обучение с учителем (supervised learning):
 - x_1, x_2, \dots, x_N - обучение без учителя
 - снижение размерности
 - кластеризация
 - $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N), x_{N+1}, x_{N+2}, \dots, x_{N+M}$ - частичное обучение (semi-supervised learning).
 - Если все новые объекты x'_1, x'_2, \dots, x'_K , для которых требуется предсказать y , известны то это трансдуктивное обучение (transductive learning).

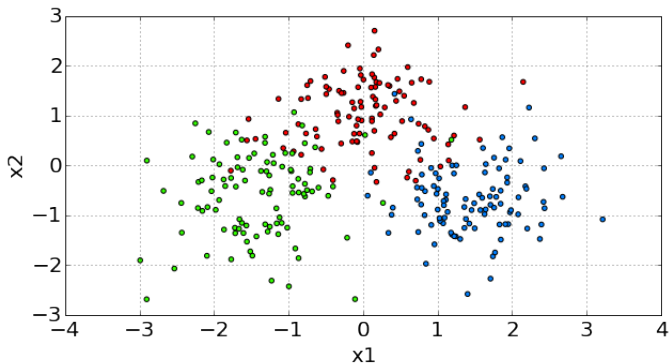
Варианты характеристик

- Тип интересующей характеристики:
 - $\mathcal{Y} = \mathbb{R}$ - регрессия (в обучении с учителем)
 - $\mathcal{Y} = \mathbb{R}^M$ - векторная регрессия (в обучении с учителем) или извлечение признаков (обучение без учителя)
 - $\mathcal{Y} = \{\omega_1, \omega_2, \dots, \omega_C\}$ - классификация (в обучении с учителем) и кластеризация (в обучении без учителя).
 - $\mathcal{Y} = \{+1, -1\}$: два класса
 - $\mathcal{Y} = \{1, 2, \dots, C\}$: C классов
 - \mathcal{Y} -множество всех подмножеств $\{\omega_1, \omega_2, \dots, \omega_C\}$ - классификация с пересекающимися классами
 - $\mathcal{Y} = \{y \in \mathbb{R}^C : y_i \in \{0, 1\}\}$, $y_i = 1 \Leftrightarrow \omega_i$ соответствует i -му объекту.

Варианты признаков

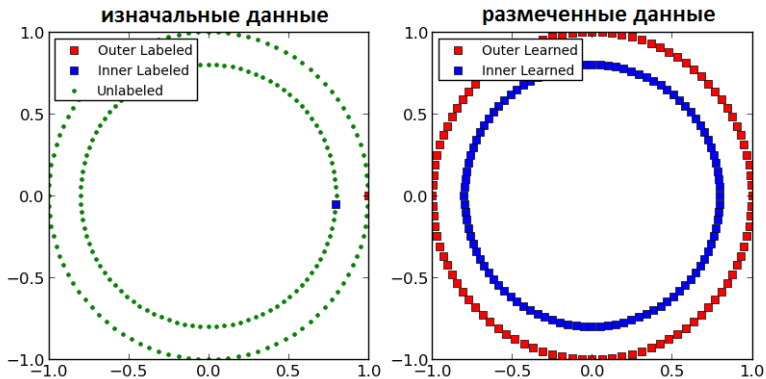
- Признаковое описание $x \in \mathcal{X}$ состоит из отдельных признаков $x_i \in \mathcal{X}_i$
- Варианты признаков
 - $\mathcal{X}_i = \{0, 1\}$ - бинарный признак
 - $|\mathcal{X}_i| < \infty$ - дискретный признак
 - $|\mathcal{X}_i| < \infty$ и \mathcal{X}_i - упорядочено - порядковый признак
 - $\mathcal{X}_i = \mathbb{R}$ - количественный признак

Пример классификации



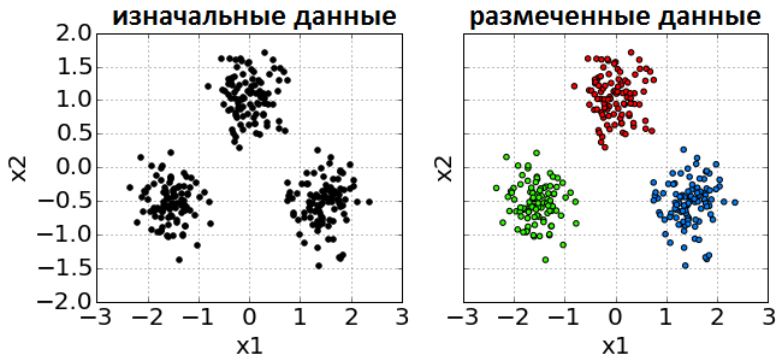
Обучение с учителем: $x = (x_1, x_2)$, y обозначен цветом

Пример частичного обучения



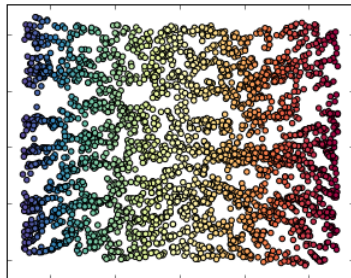
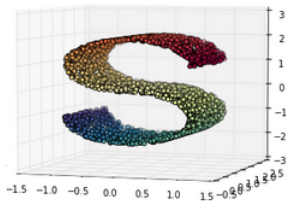
Частичное обучение.

Пример кластеризации



Обучение без учителя: кластеризация

Пример снижения размерности



Обучение без учителя: снижение размерности

Содержание

- 1 Задачи распознавания образов
- 2 Основные понятия задачи обучения с учителем.**
- 3 Практическое применение алгоритмов машинного обучения

Обучающая выборка

- **Обучающая выборка:** $X \in \mathbb{R}^{N \times D}$ - матрица объекты-признаки, $Y \in \mathbb{R}^N$ - предсказываемые величины (ответы, характеристики)
- Требуется по X восстановить параметры $\hat{\theta}$ отображения $\hat{y} = f_{\theta}(x)$ так, чтобы оно приближало истинное отображение $y = y(x)$
- Предполагается, что $z_n = (x_n, y_n)$ для $n = 1, 2, \dots, N$ - независимые одинаково распределенные случайные величины.
- Для этапа МО:
 - обучение
 - применение

Функция потерь

- **Функция потерь** $\mathcal{L}(\hat{y}, y, x)$ обычно берется $\mathcal{L}(\hat{y}, y)$
- Примеры:

- классификация:

- частота ошибок

$$\mathcal{L}(\hat{y}, y) = \mathbb{I}[\hat{y} \neq y]$$

- регрессия:

- MAE (mean absolute error):

$$\mathcal{L}(\hat{y}, y) = |\hat{y} - y|$$

- MSE (mean squared error):

$$\mathcal{L}(\hat{y}, y) = (\hat{y} - y)^2$$

- модуль относительной ошибки:

$$\mathcal{L}(\hat{y}, y) = \frac{|\hat{y} - y|}{|y|}$$

Класс функций для распознавания

- **Класс функций** - параметризованное семейство функций $F = \{f_\theta, \theta \in \Theta\}$, в котором подбирается соответствие $\mathcal{X} \rightarrow \mathcal{Y}$.
- Примеры линейного класса моделей:
 - регрессия:

$$f(x) = \theta_0 + \theta_1 x^1 + \theta_2 x^2 + \dots + \theta_D x^D,$$

где x^i - i -й признак вектора признаков x .

- классификация на два класса $y \in \{+1, -1\}$:

$$f(x) = \text{sign}\{\theta_0 + \theta_1 x^1 + \theta_2 x^2 + \dots + \theta_D x^D\},$$

где функция sign определена как

$$\text{sign}(a) = \begin{cases} 1, & a \geq 0 \\ 0, & a < 0 \end{cases}$$

Эмпирический риск

- **Алгоритм обучения** сопоставляет $f_{\hat{\theta}}(\cdot)$ по (X, Y)
 - в классе функций $F = \{f_{\theta}, \theta \in \Theta\}$
 - для заданной функции потерь $\mathcal{L}(\hat{y}, y)$
- **Эмпирический риск:**

$$L(\theta|X, Y) = \frac{1}{N} \sum_{n=1}^N \mathcal{L}(f_{\theta}(x_n), y_n)$$

- **Метод минимизации эмпирического риска:**

$$\hat{\theta} = \arg \min_{\theta} L(\theta|X, Y)$$

Оценки эмпирического риска

- Для метода минимизации эмпирического риска обычно выполнено:

$$L(\hat{\theta}|X, Y) < L(\hat{\theta}|X', Y')$$

где X, Y - выборка, на которой настраивался метод минимизации эмпирического риска, а X', Y' - новые данные.

- Оценить $L(\hat{\theta}|X', Y')$ можно с помощью:
 - контрольной выборки (не использовавшейся для оптимизации $L(\theta|X, Y)$)
 - кросс-валидации
 - метода leave-one-out

Степени обученности модели

Недообученная модель

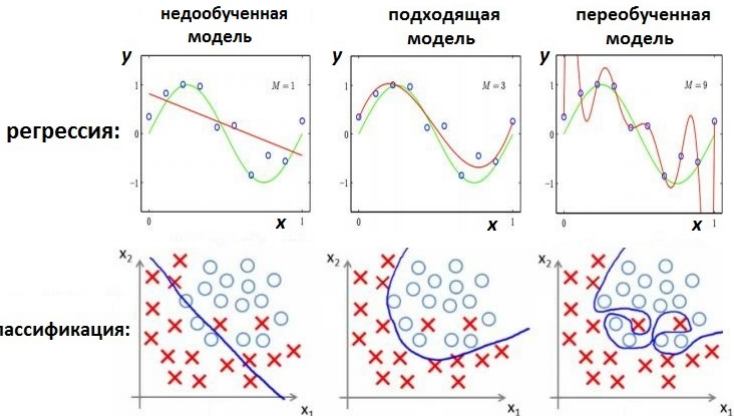
Модель, слишком сильно упрощающая закономерность $\mathcal{X} \rightarrow \mathcal{Y}$.

Переобученная модель

Модель, слишком сильно настроенная на особенности обучающей выборки (на шум в наблюдениях), а не на реальную закономерность $\mathcal{X} \rightarrow \mathcal{Y}$.

Примеры недообученных/переобученных моделей

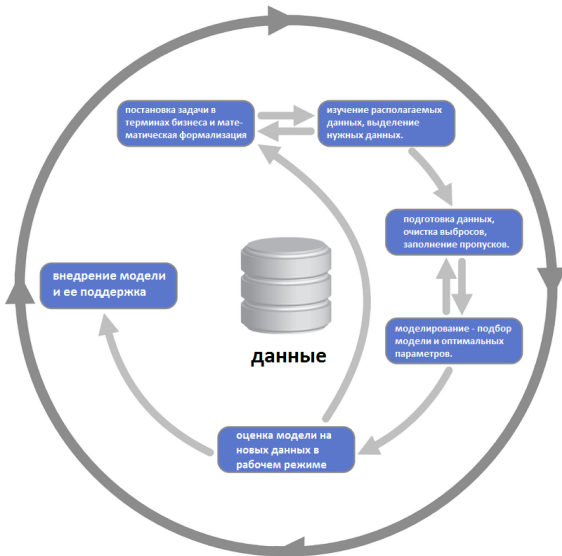
- истинная закономерность
- оцененная закономерность полиномом степени M
- объекты обучающей выборки



Содержание

- 1 Задачи распознавания образов
- 2 Основные понятия задачи обучения с учителем.
- 3 Практическое применение алгоритмов машинного обучения

Методология CrispDM



Примеры прикладных задач

Классификация:

- сопоставление тематических рубрик текстовым документам
- в поисковой системе - соответствуют ли друг другу поисковый запрос и документ?
- является ли последовательность транзакций в сети регулярной последовательностью или попыткой взлома системы?
- уйдет ли клиент с заданной историей звонков, платежей и используемых услуг к другому мобильному оператору?
- погасит ли клиент банка взятый кредит в полном объеме?
- свидетельствует ли полученный сигнал от радара о наличии самолета, или это шум?

Регрессия:

- определить цену на квартиру по ее характеристикам
- определить величину спроса в магазине за некоторый будущий период

Система обозначений, используемых в курсе

- Если это соответствует контексту изложения, и нет переобозначений, то:
 - x - вектор признакового описания объекта
 - y - оцениваемая характеристика объекта с вектором признаков x
 - x_i - i -й объект обучающей выборки, y_i - соответствующая характеристика
 - x^k - k -й признак признакового описания x
 - x_i^k - k -й признак признакового описания x_i
 - D - размерность признакового пространства: $x \in \mathbb{R}^D$
 - N - количество объектов в обучающей выборке
 - X - матрица объекты-признаки, $X \in \mathbb{R}^{N \times D}$
 - $Y \in \mathbb{R}^N$ - вектор характеристик объектов обучающей выборки
 - $\mathcal{L}(\hat{y}, y)$ - функция цены при предсказании истинного значения y значением \hat{y} .
 - в задачах классификации $\{\omega_1, \omega_2, \dots, \omega_C\}$ - множество классов, C - общее количество классов.
 - \hat{z} обозначает оценку z по данным обучающей выборки: например, $\hat{\theta}$ - оценка θ , \hat{y} - оценка y и т.д.