

# Лекция 1. Вероятностные модели в задачах машинного обучения

Д. П. Ветров<sup>1</sup>    Д. А. Кропотов<sup>2</sup>

<sup>1</sup>МГУ, ВМиК, каф. ММП

<sup>2</sup>ВЦ РАН

Спецкурс «Байесовские методы машинного обучения»

# Цели курса

## Лекция 1.

Вероятностные модели в задачах машинного обучения

Ветров,  
Кропотов

Ликбез

Некоторые задачи машинного обучения

Вероятностные модели

Основная проблема машинного обучения

- Ознакомление с вероятностными моделями обработки данных, особенностями их применения на практике и их недостатками
- Описание вероятностных моделей для нескольких практически важных задач анализа данных (отбор информативных признаков, разделение смесей распределений, уменьшение размерности данных и др.)
- Изложение современных достижений в области практического использования байесовских методов
- Развитие навыков построения вероятностных моделей и работы с ними
- Напоминание основных результатов из смежных дисциплин (теория кодирования, анализ, матричные вычисления, статистика, линейная алгебра, теория вероятностей)

# План лекции

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

## 1 Ликбез

Основные понятия мат. статистики

## 2 Некоторые задачи машинного обучения

Задача классификации

Задача восстановления регрессии

Задача кластеризации (обучения без учителя)

Задача идентификации

Задача уменьшения размерности

## 3 Вероятностные модели

## 4 Основная проблема машинного обучения

# Краткое напоминание основных вероятностных понятий

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Основные  
понятия мат.  
статистики

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- $X : \Omega \rightarrow \mathbb{R}$  – случайная величина
- Вероятность попадания величины в интервал  $(a, b)$  равна

$$P(a \leq X \leq b) = \int_a^b p(x)dx,$$

где  $p(x)$  – плотность распределения  $X$ ,

$$p(x) \geq 0, \quad \int_{-\infty}^{\infty} p(x)dx = 1$$

- Если поведение случайной величины определяется некоторым параметром, возникают условные плотности  $p(x|\theta)$ . Если рассматривать условную плотность как функцию от параметра

$$f(\theta) = p(x|\theta),$$

то принято говорить о т.н. функции правдоподобия

# Основная задача мат. статистики

Лекция 1.

Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Основные  
понятия мат.  
статистики

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Распределение случайной величины  $X$  известно с точностью до параметра  $\theta$
- Имеется выборка значений величины  $X$ ,  $\mathbf{x} = (x_1, \dots, x_n)$
- Требуется оценить значение  $\theta$
- Метод максимального правдоподобия

$$\hat{\theta}_{ML} = \arg \max_{\theta} f(\theta) = \arg \max_{\theta} p(\mathbf{x}|\theta) = \arg \max_{\theta} \prod_{i=1}^n p(x_i|\theta)$$

- Можно показать, что ММП является асимптотически оптимальным при  $n \rightarrow \infty$
- Увы, мир несовершенен. Величина  $n$  конечна и обычно не слишком велика

# Пример использования

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез  
Основные  
понятия мат.  
статистики

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Пусть имеется выборка из нормального распределения  $\mathcal{N}(x|\mu, \sigma^2)$  с неизвестными мат. ожиданием и дисперсией
- Выписываем логарифм функции правдоподобия

$$L(X|\mu, \sigma) = - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} - n \log \sigma - \frac{n}{2} \log(2\pi) \rightarrow \max_{\mu, \sigma}$$

$$\frac{\partial L}{\partial \mu} = - \sum_{i=1}^n \frac{(x_i - \mu)}{\sigma^2} = 0 \quad \frac{\partial L}{\partial \sigma} = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma} = 0$$

$$\begin{aligned} \mu_{ML} &= \frac{1}{n} \sum_{i=1}^n x_i & \sigma_{ML}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \\ & & &= \frac{1}{n} \sum_{i=1}^n \left( x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2 \end{aligned}$$

# Условная вероятность

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Основные  
понятия мат.  
статистики

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Пусть  $X$  и  $Y$  — случайные величины с плотностями  $p(x)$  и  $p(y)$  соответственно
- В общем случае их совместная плотность  $p(x, y) \neq p(x)p(y)$ . Если это равенство выполняется, величины называют **независимыми**
- Условной плотностью называется величина

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

- Смысл: как факт  $Y = y$  влияет на распределение  $X$ . Заметим, что  $\int p(x|y)dx \equiv 1$ , но  $\int p(x|y)dy$  не обязан равняться единице, т.к. относительно  $y$  это не плотность, а **функция правдоподобия**
- Очевидная система тождеств  $p(x|y)p(y) = p(x, y) = p(y|x)p(x)$  позволяет легко переходить от  $p(x|y)$  к  $p(y|x)$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

# Концепция машинного обучения

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации  
Задача  
восстановления  
регрессии  
Задача  
кластеризации  
(обучения без  
учителя)  
Задача  
идентификации  
Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Большинство математических задач сводятся к определению значений некоторых величин по заданному набору других величин
- Предполагается, что между величинами существует зависимость
- Выделяют два подхода к решению таких задач: математическое моделирование (model-based reasoning) и восстановление зависимостей путем обработки прошлого опыта
- Основное требование для применимости второго подхода – наличие обучающей информации
- Как правило в качестве таковой выступает выборка **прецедентов** – ситуационных примеров из прошлого



# Структура прецедента

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации  
Задача  
восстановления  
регрессии  
Задача  
кластеризации  
(обучения без  
учителя)  
Задача  
идентификации  
Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Атомарным элементом информации является прецедент — объект, обладающий набором наблюдаемых и скрытых переменных
- В обучающей выборке значения скрытых переменных известны
- Требуется построить алгоритм, который позволял бы оценить зависимость между наблюдаемыми и скрытыми переменными объекта по обучающей выборке и использовать ее для обработки новых, не встречавшихся ранее объектов, значения скрытых переменных которых неизвестны

# Классификация

Лекция 1.

Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации

Задача  
восстановления  
регрессии

Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Исторически возникла из задачи машинного зрения, поэтому часто употребляемый синоним – распознавание образов
- В классической задаче классификации обучающая выборка представляет собой набор отдельных объектов  $(X, T) = \{(\mathbf{x}_i, t_i)\}_{i=1}^n$
- У каждого объекта есть наблюдаемые переменные (признаки)  $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,d}) \in \mathbb{R}^d$  и скрытая переменная  $t_i$ , принимающая конечное число значений, обычно из множества  $\mathcal{T} = \{1, \dots, l\}$
- Требуется построить алгоритм (классификатор), который по вектору признаков  $\mathbf{x}$  вернул бы метку класса  $\hat{t}$  или вектор оценок принадлежности (апостериорных вероятностей) к каждому из классов  $\{p(t|\mathbf{x})\}_{t=1}^l$

# Классификация

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации

Задача  
восстановления  
регрессии

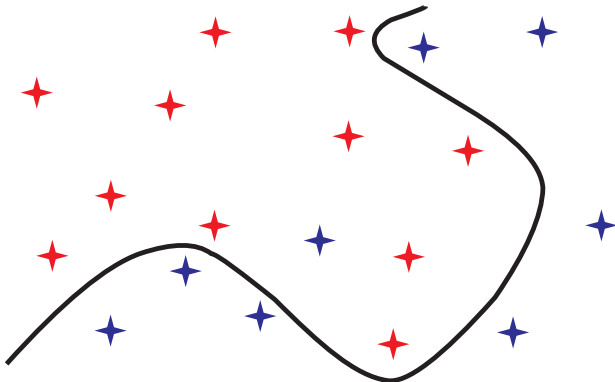
Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения



# Примеры задач классификации

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации

Задача  
восстановления  
регрессии

Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Кредитный скоринг: по анкете заемщика принять решение о выдаче/отказе кредита
- Социология: по анкете требуется отнести респондента к одной из фокус-групп
- Оптическое распознавание символов: по отсканированному изображению символа определить его тип
- Синтез химических соединений: по параметрам химических элементов определить, будет ли получаемые соединения обладать заданным свойством или нет

# Регрессия

## Лекция 1.

Вероятностные модели в задачах машинного обучения

Ветров,  
Кропотов

Ликбез

Некоторые задачи машинного обучения

Задача классификации

Задача восстановления регрессии

Задача кластеризации (обучения без учителя)

Задача идентификации

Задача уменьшения размерности

Вероятностные модели

Основная проблема машинного обучения

- Исторически возникла при исследовании влияния одной группы непрерывных случайных величин на другую группу непрерывных случайных величин
- В классической задаче восстановления регрессии обучающая выборка представляет собой набор отдельных объектов  $(X, T) = \{(\mathbf{x}_i, t_i)\}_{i=1}^n$
- У каждого объекта есть наблюдаемые переменные (признаки)  $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,d}) \in \mathbb{R}^d$  и скрытая переменная  $t_i \in \mathbb{R}$
- Требуется построить алгоритм (регрессор), который по вектору признаков  $\mathbf{x}$  вернул бы точечную оценку значения регрессии  $\hat{t}$ , доверительный интервал  $(t_-, t_+)$  или апостериорное распределение на множестве значений скрытой переменной  $p(t|\mathbf{x})$

# Регрессия

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации

Задача  
восстановления  
регрессии

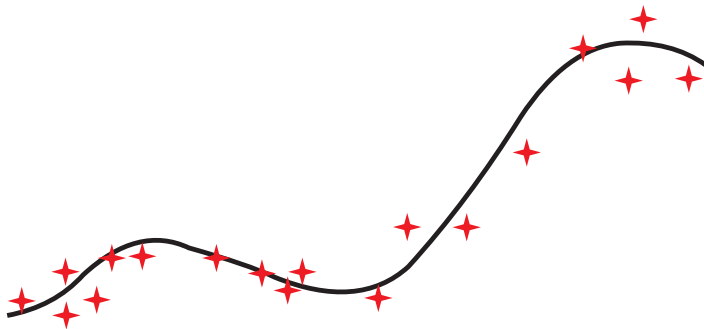
Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения



# Примеры задач восстановления регрессии

## Лекция 1.

Вероятностные модели в задачах машинного обучения

Ветров,  
Кропотов

Ликбез

Некоторые задачи машинного обучения

Задача классификации

Задача восстановления регрессии

Задача кластеризации (обучения без учителя)

Задача идентификации

Задача уменьшения размерности

Вероятностные модели

Основная проблема машинного обучения

- Оценка стоимости недвижимости: по характеристике района, экологической обстановке, транспортной связности оценить стоимость жилья
- Прогноз свойств соединений: по параметрам химических элементов спрогнозировать температуру плавления, электропроводность, теплоемкость получаемого соединения
- Кредитный скоринг: по анкете заемщика оценить величину кредитного лимита
- Инженерное дело: по техническим характеристикам автомобиля и режиму езды спрогнозировать расход топлива

# Кластеризация

## Лекция 1.

Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации  
Задача  
восстановления  
регрессии

Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации  
Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Исторически возникла из задачи группировки схожих объектов в единую структуру (кластер) с последующим выявлением общих черт
- В классической задаче кластеризации обучающая выборка представляет собой набор отдельных объектов  $X = \{\mathbf{x}_i\}_{i=1}^n$ , характеризующихся вектором наблюдаемых переменных (признаков)  
$$\mathbf{x}_i = (x_{i,1}, \dots, x_{i,d}) \in \mathbb{R}^d$$
- Требуется построить алгоритм (кластеризатор), который разбил бы выборку на непересекающиеся группы (кластеры)  $X = \bigcup_{j=1}^k C_k$ ,  $C_j \subset \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ ,  $C_i \cap C_j = \emptyset$
- В каждый класс должны попасть объекты в некотором смысле похожие друг на друга
- Ясно, что такую задачу можно сформулировать в терминах скрытых переменных (Упр.)



# Кластеризация

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

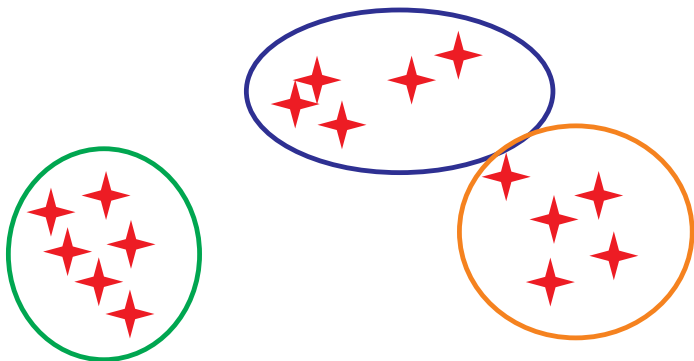
Задача  
классификации  
Задача  
восстановления  
регрессии

Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации  
Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения



# Примеры задач кластерного анализа

## Лекция 1.

Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации

Задача  
восстановления  
регрессии

Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Экономическая география: по физико-географическим и экономическим показателям разбить страны мира на группы схожих по экономическому положению государств
- Финансовая сфера: по сводкам банковских операций выявить группы «подозрительных», нетипичных банков, сгруппировать остальные по степени близости проводимой стратегии
- Маркетинг: по результатам маркетинговых исследований среди множества потребителей выделить характерные группы по степени интереса к продвигаемому продукту
- Социология: по результатам социологических опросов выявить группы общественных проблем, вызывающих схожую реакцию у общества, а также характерные фокус-группы населения

# Идентификация

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации  
Задача  
восстановления  
регрессии  
Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Исторически возникла из классификации, необходимости отделить объекты, обладающие определенным свойством, от «всего остального»
- В классической задаче идентификации обучающая выборка представляет собой набор отдельных объектов  $(X, T) = \{(\mathbf{x}_i, t_i)\}_{i=1}^n$
- У каждого объекта есть наблюдаемые переменные (признаки)  $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,d}) \in \mathbb{R}^d$  и скрытая переменная  $t_i \in \{0, 1\}$ , отражающая наличие желаемого свойства
- Для всех объектов обучающей выборки известно, что  $t_i = 1$ , при этом не существует возможности сделать репрезентативную выборку для объектов с  $t = 0$
- Требуется построить алгоритм (идентификатор), который по вектору признаков  $\mathbf{x}$  определил бы значение скрытой переменной  $t$ , либо вернул распределение вероятностей на ее значения  $p(t|\mathbf{x})$

# Идентификация

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

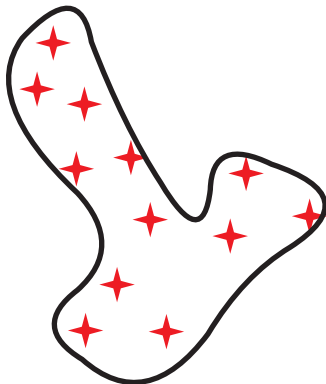
Задача  
классификации  
Задача  
восстановления  
регрессии  
Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения



# Примеры задач идентификации

## Лекция 1.

Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации  
Задача  
восстановления  
регрессии  
Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Медицинская диагностика: по набору медицинских характеристик требуется установить наличие/отсутствие конкретного заболевания
- Системы безопасности: по камерам наблюдения в подъезде идентифицировать жильца дома
- Банковское дело: определить подлинность подписи на чеке
- Обработка изображений: выделить участки с изображениями лиц на фотографии
- Искусствоведение: по характеристикам произведения (картины, музыки, текста) определить, является ли его автором тот или иной мастер

# Уменьшение размерности

## Лекция 1.

Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации

Задача  
восстановления  
регрессии

Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Объекты, описываемые  $d$ -мерным вектором признаков часто лежат в подпространстве (линейном или нелинейном) меньшей размерности
- В классической задаче уменьшения размерности обучающая выборка представляет собой набор отдельных объектов  $X = \{\mathbf{x}_i\}_{i=1}^n$ , характеризующихся вектором наблюдаемых переменных (признаков)  $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,d}) \in \mathbb{R}^d$
- Требуется построить алгоритм, осуществляющий преобразование  $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ , где  $p < d$
- Взаиморасположения объектов, задаваемых новыми векторами признаков  $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$  должны в наибольшей степени соответствовать взаиморасположениям объектов  $\mathbf{x}_1, \dots, \mathbf{x}_n$
- Ясно, что такую задачу можно сформулировать в терминах скрытых переменных (Упр.)

# Уменьшение размерности

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Задача  
классификации

Задача  
восстановления  
регрессии

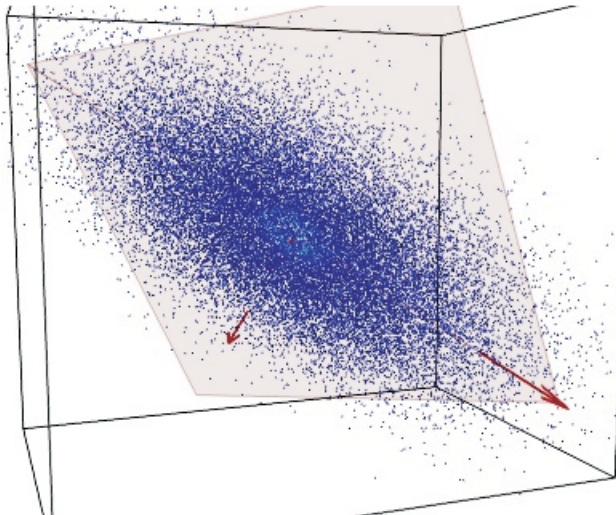
Задача  
кластеризации  
(обучения без  
учителя)

Задача  
идентификации

Задача  
уменьшения  
размерности

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения



# Примеры задач уменьшения размерности

## Лекция 1.

Вероятностные модели в задачах машинного обучения

Ветров,  
Кропотов

Ликбез

Некоторые задачи машинного обучения

Задача классификации

Задача восстановления регрессии

Задача кластеризации (обучения без учителя)

Задача идентификации

Задача уменьшения размерности

Вероятностные модели

Основная проблема машинного обучения

- Визуализация многомерных данных: найти двухмерное представление объектов, в минимальной степени искажающее взаимные расстояния
- Составление фотороботов: выделить параметры, отвечающие за характерные вариации лиц
- Математическое моделирование: построение статистических моделей сложных многомерных объектов
- Анализ текстов: по фрагменту текста определить его тематику/рубрику
- Маркетинг: выделить факторы, влияющие на поведение покупателя



# Общая формулировка задач обработки данных

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

Все перечисленные выше задачи (и многие другие) обладают рядом общих черт

- Имеется массив объектов с наблюдаемыми переменными  $X$  и скрытыми переменными  $T$
- Предполагается, что между наблюдаемыми и скрытыми переменными существует зависимость
- Точный вид этой зависимости нам неизвестен и/или зависимость недетерминированная, т.е. значения наблюдаемых переменных не позволяют однозначно определить значения скрытых переменных

При вероятностном подходе к решению этих задач, неопределенность в зависимости между  $X$  и  $T$  моделируется введением совместного распределения на все переменные  $p(X, T)$ . Выделяют два вида вероятностных моделей: порождающие (generative) и дискриминативные (discriminative)

# Порождающая модель

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- При использовании порождающих моделей необходимо задать совместное распределение  $p(X, T)$  на множестве объектов
- Зная совместное распределение мы можем моделировать новые объекты из той же генеральной совокупности
- Если совместное распределение неизвестно (т.е. почти всегда), можно пробовать его настроить по обучающей выборке
- Как правило, это чрезвычайно сложная задача, решение которой, к тому же, избыточно в большинстве задач

# Дискриминативная модель

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- При использовании дискриминативных моделей необходимо знать условное распределение  $p(T|X)$  на множестве значений скрытых переменных объекта
- Зная условное распределение мы можем определить наиболее вероятные значения скрытых переменных объекта
- В отличие от порождающей модели, дискриминативная модель не позволяет моделировать новые объекты из генеральной совокупности.
- В случае, когда условная плотность неизвестна, ее можно попробовать настроить по обучающей выборке
- Настройка дискриминативной модели более простая, поэтому если нам требуется только уметь определять значения скрытых переменных по наблюдаемым, использование такой модели предпочтительно

# Основные задачи, решаемые с помощью вероятностных моделей

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Обучение с учителем. Дана обучающая выборка  $(X, T)$  и параметрически задано распределение  $p(X, T|\theta)$ . Задача определить значение параметров  $\theta$ , при которых распределение наилучшим образом описывает обучающую выборку  $p(X, T|\theta) \rightarrow \max_{\theta}$
- Обучение без учителя. Дана выборка объектов с наблюдаемыми переменными  $X$  и параметрически задано распределение  $p(X, T|\theta)$ . Задача определить значение параметров  $\theta$ , при которых распределение наилучшим образом описывает наблюдаемые данные  $p(X|\theta) \rightarrow \max_{\theta}$
- MAP-оценивание. Дана вероятностная модель  $p(X, T)$  и массив объектов с известными наблюдаемыми переменными  $X$ . Задача определить наиболее вероятные значения скрытых переменных объектов  $p(X, T) \rightarrow \max_T$
- Оценка маргиналов. Дана вероятностная модель  $p(X, T)$  и массив объектов с известными наблюдаемыми переменными  $X$ . Задача определить маргинальное распределение на отдельно взятую скрытую переменную  $p(t|X) - ?$

Аналогичные задачи могут быть сформулированы и для дискриминативных моделей

# Байесовские методы работы с вероятностными моделями

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Для работы с вероятностными моделями обычно используются т.н. байесовские методы, базирующиеся на одноименном подходе к теории вероятностей (следующая лекция)
- Основным достоинством байесовской парадигмы является возможность учета наших предпочтений на вид модели, характерные значения скрытых переменных, желаемые свойства оцениваемых параметров и пр.
- Как это ни удивительно, при решении практических задач таких предпочтений оказывается довольно много, хотя они не всегда лежат на поверхности
- Байесовский аппарат предоставляет удобное средство строгой формализации таких предпочтений

# Многомерные скрытые переменные

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- До сих пор предполагалось, что множество наблюдаемых переменных у всех объектов одинаково, а скрытая переменная скалярная
- Формализм вероятностных моделей может быть применен (и успешно применяется!) и для более сложных случаев
- В частности, особый интерес представляют ситуации, когда скрытые переменные объекта многомерные и связаны сложными взаимозависимостями
- Примерами таких объектов являются социальные сети, изображения, сигналы и др.
- Обучающая выборка при этом может состоять из одного сложного объекта!
- Методы работы с такими объектами рассматриваются в курсе «Графические модели» (на правах рекламы ;-))

# Настройка параметров на обучающую выборку

Лекция 1.

Вероятностные модели в задачах машинного обучения

Ветров,  
Кропотов

Ликбез

Некоторые задачи машинного обучения

Вероятностные модели

Основная проблема машинного обучения

- Рассмотрим порождающую модель  $p(X, T|\theta)$ , заданную с точностью до параметров  $\theta$
- Прямой способ обучения по обучающей выборке  $(X, T)$  предполагает максимизацию функции  $p(X, T|\theta)$  по параметрам  $\theta$ , т.е. применение метода максимального правдоподобия
- Аналогично в нестатистических методах машинного обучения предполагается настройка параметров путем оптимизации некоторого функционала качества, определенного на обучающей выборке
- Но разве это то, что нужно?.. Ведь обучающая выборка конечна и может содержать (и почти всегда содержит!) помехи

# Проблема переобучения

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

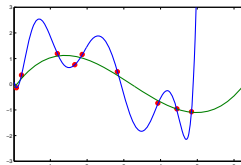
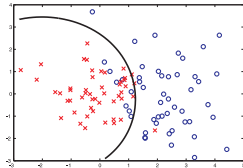
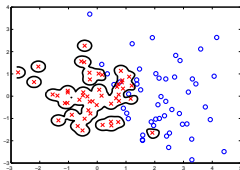
Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

Прямая максимизация качества на обучающей выборке без ограничений на настраиваемые параметры ведет к получению решающих правил, способных объяснить все что угодно и найти закономерности даже там, где их нет.





# Способы оценки обобщающей способности

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- На сегодняшний день единственным универсальным способом оценивания обобщающей способности является кросс-валидация
- Все попытки предложить что-нибудь отличное от метода проб и ошибок пока **не привели к общепризнанному решению**. Наиболее известны из них следующие:
  - Структурная минимизация риска (В. Вапник, А. Червоненкис, 1974)
  - Минимизация длины описания (Дж. Риссанен, 1978)
  - Информационные критерии Акаике и Байеса-Шварца (Акаике, 1974, Шварц, 1978)
  - Байесовская регуляризация (МакКай, 1992)
- Последний принцип трактует параметры  $\theta$  как случайные величины и добавляет их в вероятностную модель  $p(X, T, \theta)$ . Вопрос определения конкретного вида нашей вероятностной модели известен как задача выбора модели

# Байесовская регуляризация

Лекция 1.

Вероятностные модели в задачах машинного обучения

Ветров,  
Кропотов

Ликбез

Некоторые задачи машинного обучения

Вероятностные модели

Основная проблема машинного обучения

- Параметры  $\theta$  стали частью вероятностной модели
- Совместное распределение  $p(X, T, \theta)$  может быть представлено в следующем виде

$$p(X, T, \theta) = p(X, T|\theta)p(\theta)$$

- У нас появился множитель  $p(\theta)$ , с помощью которого можно ограничить множество допустимых значений  $\theta$  и/или ввести предпочтения на те или иные значения
- Учитывая, что функция  $p(X, T|\theta)$  нам была известна, задача выбора модели свелась к определению априорного распределения  $p(\theta)$
- Введение ограничений на возможные значения настраиваемых в ходе обучения параметров, не индуцированных обучающей выборкой часто называют **регуляризацией**

# Примеры задач выбора модели

Лекция 1.  
Вероятностные  
модели в задачах  
машинного  
обучения

Ветров,  
Кропотов

Ликбез

Некоторые  
задачи  
машинного  
обучения

Вероятностные  
модели

Основная  
проблема  
машинного  
обучения

- Определение числа кластеров в данных
- Выбор коэффициента регуляризации в задаче машинного обучения (например, коэффициента затухания весов (weight decay) в нейронных сетях)
- Установка степени полинома при интерполяции сплайнами
- Выбор наилучшей ядерной функции в методе опорных векторов (SVM)
- Определение количества ветвей в решающем дереве
- и многое другое...