ТЕОРЕМА БАЙЕСА

МФТИ

УСЛОВНАЯ ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ

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

$$p(y|x)p(x) = p(x,y) = p(x|y)p(y) \;\Rightarrow$$

Формула обращения условной плотности

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

Условная плотность распределения

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

$$1 = \int p(x|y)dx = \int \frac{p(y|x)p(x)}{p(y)}\, dx =$$

$$= \frac{1}{p(y)} \int p(y|x)p(x)dx$$

Отсюда $p(y) = \int p(y|x)p(x)dx$ — маргинализация или правило суммирования

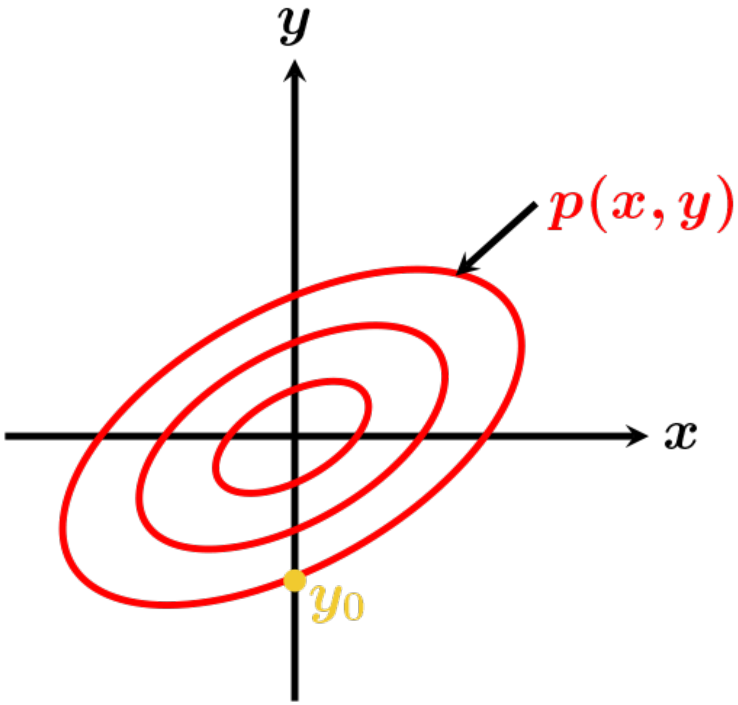
МФТИ

ТЕОРЕМА БАЙЕСА

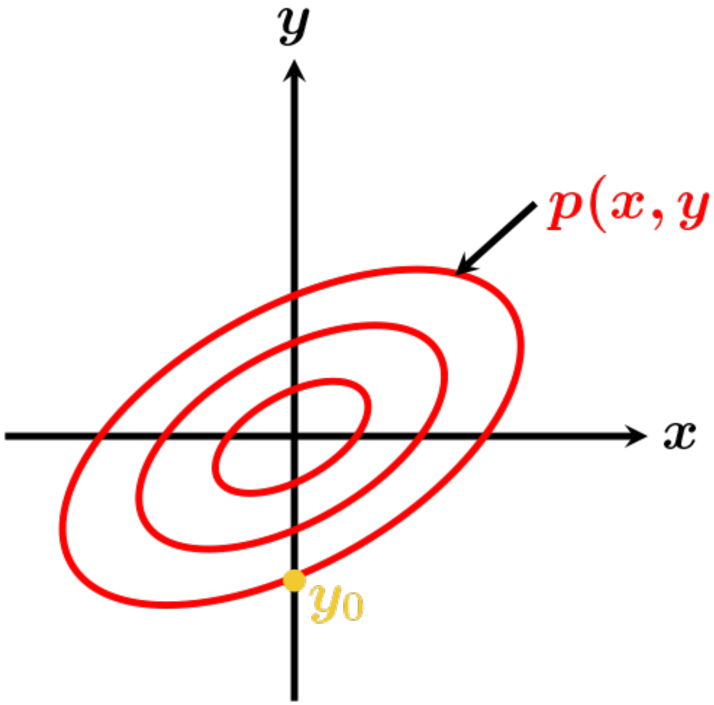
$$p(y|x) = \frac{p(x|y)p(y)}{\int p(x|y)p(y)dy}$$

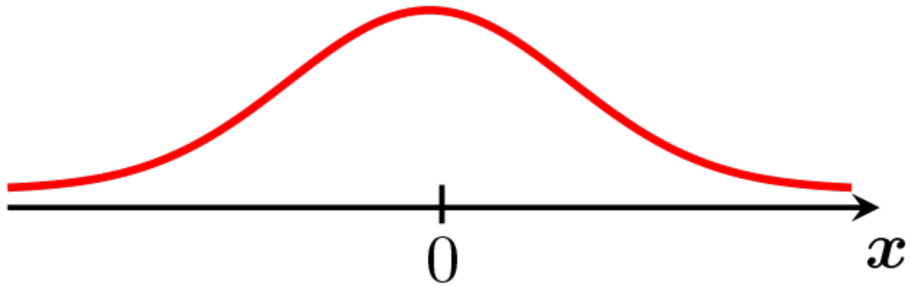
$$\textbf{Posterior} = \frac{\textbf{Likehood} \times \textbf{Prior}}{\textbf{Evidence}}$$

Условное и маргинальное распределения



Условное и маргинальное распределения

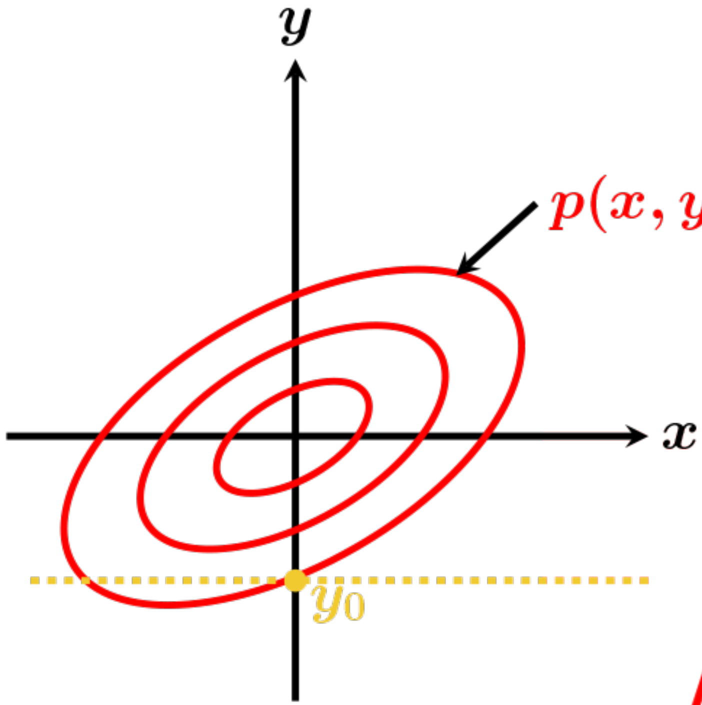




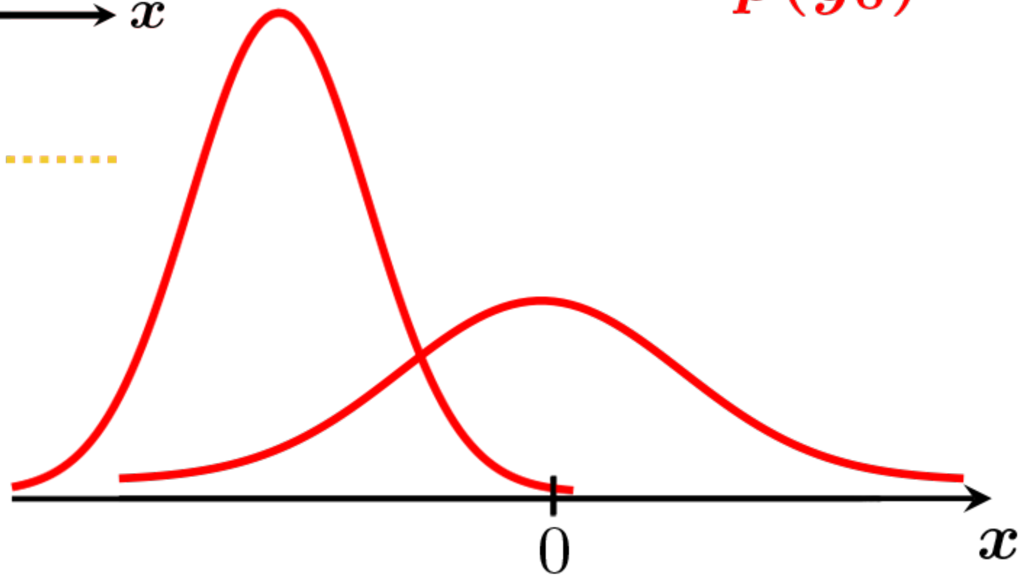
$$p(x) = \int p(x, y)dy = \int p(x|y)p(y)dy$$

МФТИ

Условное и маргинальное распределения



$$p(x|y_0) = \frac{p(x, y_0)}{p(y_0)}$$



$$p(x) = \int p(x, y)dy = \int p(x|y)p(y)dy$$

Метод максимального правдоподобия

- Пусть $X = (x_1, \ldots, x_n)$ — выборка из н.о.р.с.в. $x_i \sim p(x|\theta)$
- $p(x|\theta)$ — плотность распределения, известная с точностью до параметров θ
- Задача: оценить θ по выборке X
- Метод максимального правдоподобия:

$$\theta_{ML} = \arg\max_{\theta} p(X|\theta) = \arg\max_{\theta} \prod_{i=1}^{n} p(x_i|\theta) = \\ = \arg\max_{\theta} \sum_{i=1}^{n} \log p(x_i, \theta)$$

МФТИ

МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

- Достоинства метода:
  - Ассимптотическая несмещённость, т.е. $\mathbb{E}\,\theta_{ML} = \theta_*$ при $n \gg 1$
  - Состоятельность: $\theta_{ML} \to \theta_*$ при $n \to +\infty$
  - Ассимптотическая нормальность
  - Эффективность

- Недостаток:
  - Основные теоретические гарантии получены при $\frac{n}{d} \gg 1$, где d — размерность θ

МФТИ

РЕЗЮМЕ

- Условное и безусловное распределения
- Теорема Байеса
- Метод максимального правдоподобия

БАЙЕСОВСКИЙ ПОДХОД К ТЕОРИИ ВЕРОЯТНОСТЕЙ

ЧАСТОТНЫЙ И БАЙЕСОВСКИЙ ПОДХОДЫ

	Частотный	Байесовский
Интерпертация случайности	Объективная неопределённость	Субъективное незнание
Метод вывода	Метод максимального правдоподобия	Теорема Байеса
Оценки	$\boldsymbol{\theta_{ML}}$	$\boldsymbol{p(\theta \mid X)}$
Применимость	$\boldsymbol{\frac{n}{d} \gg 1}$	$\forall\ \boldsymbol{n}$

МФТИ

ИЛЛЮСТРАТИВНЫЙ ПРИМЕР

- Имеется несколько косвенных проявлений $y_1, \ldots, y_m$ неизвестной величины x
- Для каждой из них существует вероятностная модель, определяющая вероятность наблюдения того или иного значения y_j : $p_j(y_j|x)$
- Задача: оценить x путём объединения информации о всех $y_1, \ldots, y_m$

МФТИ

ИЛЛЮСТРАТИВНЫЙ ПРИМЕР

- Фиксируем исходные представления о возможных значениях x в виде $p(x)$
- Выполняем байесовский вывод относительно y_1:

$$p(x|y_1) = \frac{p_1(y_1|x)p(x)}{\int p_1(y_1|x)p(x)dx}$$

МФТИ

ИЛЛЮСТРАТИВНЫЙ ПРИМЕР

- Выполняем байесовский вывод относительно y_2, используя результат предыдущего шага в качестве нового априорного распределения:

$$p(x|y_1, y_2) = \frac{p_2(y_2|x)p(x|y_1)}{\int p_2(y_2|x)p(x|y_1)dx}$$

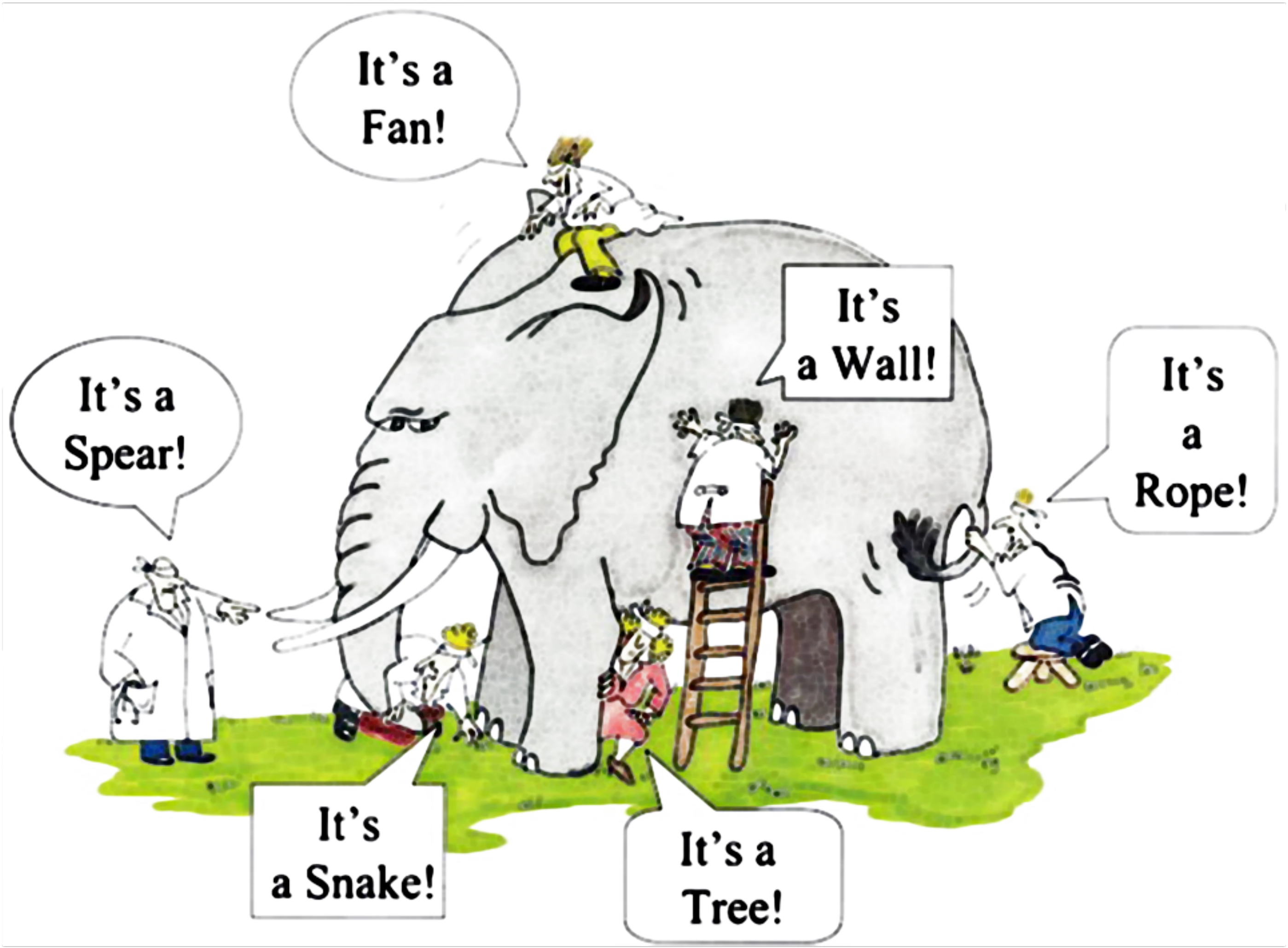
- ...

- Считаем $p(x|y_1, \ldots, y_m) =$

$$= \frac{p_m(y_m|x)p(x|y_1, \ldots, y_{m-1})}{\int p_m(y_m|x)p(x|y_1, \ldots, y_{m-1})dx}$$

МФТИ

ИЛЛЮСТРАТИВНЫЙ ПРИМЕР



РЕЗЮМЕ

- Байесовский подход к теории вероятностей
- Объединение нескольких вероятностных моделей в более сложную модель

БАЙЕСОВСКИЕ МОДЕЛИ В ЗАДАЧАХ МАШИННОГО ОБУЧЕНИЯ

БАЙЕСОВСКИЙ ПОДХОД

- При использовании подхода нам необходимо задать совместное распределение на все переменные, которые мы хотим предсказывать на этапах обучения или тестирования

ПРИМЕР: ЛИНЕЙНАЯ РЕГРЕССИЯ

- Рассмотрим задачу линейной регрессии:
  $x \in \mathbb{R}^d$, $\mathbf{w} \in \mathbb{R}^d$, $t \in \mathbb{R}$

- Три группы переменных:
  - x — признаки
  - t — целевая переменная
  - $\mathbf{w}$ — веса линейной регрессии

- Вероятностная модель:

$$p(t, \mathbf{w} \,|x) = p(t|x, \mathbf{w})p(\mathbf{w}) = $$
$$= \mathcal{N}(t|\,\mathbf{w}^T x, \sigma^2)\mathcal{N}(\mathbf{w}\,|0, I)$$

ПРИМЕР: ЛИНЕЙНАЯ РЕГРЕССИЯ

- Пусть задана обучающая выборка

$$(\boldsymbol{X}, \boldsymbol{T}) = (\boldsymbol{x}_i, \boldsymbol{t}_i)_{i=1}^{n},$$

тогда

ПРИМЕР: ЛИНЕЙНАЯ РЕГРЕССИЯ

$$\mathrm{w}_{MP} = \operatorname*{argmax}_{\mathrm{w}} p(\mathrm{w}\,|X,T) = \operatorname*{argmax}_{\mathrm{w}} p(T|X,\mathrm{w})p(\mathrm{w}) =$$

$$= \operatorname*{argmax}_{\mathrm{w}} \prod_{i=1}^{n} p(t_i|x_i,\mathrm{w})p(\mathrm{w}) =$$

$$= \operatorname*{argmax}_{\mathrm{w}} \left[\sum_{i=1}^{n} \log p(t_i|x_i,\mathrm{w}) + \log p(\mathrm{w})\right] =$$

$$= \operatorname*{argmin}_{\mathrm{w}} \frac{1}{2\sigma^2} \sum_{i=1}^{n} (t_i - x_i^T\,\mathrm{w})^2 + \frac{1}{2}\|\,\mathrm{w}\,\|^2 =$$

$$= \operatorname*{argmin}_{\mathrm{w}} \left[\sum_{i=1}^{n} (t_i - x_i^T\,\mathrm{w})^2 + \sigma^2\|\,\mathrm{w}\,\|^2\right]$$

Поиск максимума апостериорной плотности эквивалентен МНК с L_2-регуляризатором

МФТИ

ЗАЧЕМ НУЖНЫ БАЙЕСОВСКИЕ МЕТОДЫ?

- Возможность строить сложные модели из простых путём использования выходов (апостериорного распределения) одной модели в качестве входов (априорного распределения) в другой

ЗАЧЕМ НУЖНЫ БАЙЕСОВСКИЕ МЕТОДЫ?

- Возможность обработки последовательно поступающих данных “на лету” без необходимости повторного обучения с нуля

ЗАЧЕМ НУЖНЫ БАЙЕСОВСКИЕ МЕТОДЫ?

- Возможность вводить на настраиваемые параметры априорные распределения, которые:
  - отражают наши предпочтения о значениях параметров
  - предотвращают переобучение

ЗАЧЕМ НУЖНЫ БАЙЕСОВСКИЕ МЕТОДЫ?

- Возможность обучаться по не полностью размеченным , частично размеченным и неразмеченным обучающим выборкам

Резюме

- Применение байесовского метода к линейной регрессии
- Преимущества байесовских методов