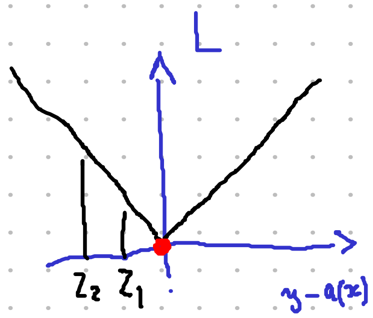
а) $\exists\, L(y, a(x)) = g(y - a(x))$. $|y - a(x)|$

б) $g(0) = 0$

→ в) $|z_1| < |z_2| \Rightarrow g(z_1) \le g(z_2)$

г) g — гладкий



$$L(y, a(x)) = g(y - a(x)) \simeq$$

$$\simeq \underbrace{g(0)}_{0} + \underbrace{g'(0)}_{0} \cdot (y - a(x)) + \frac{g''(0)}{2}(y - a(x))^2 =$$

$$= c \cdot (y - a(x))^2$$

Вероятностный подход: ММП $\mapsto$ ф. ошибки

$$(-1)\cdot \log \underline{\mathbb{P}(y \mid x, \theta)} \longrightarrow \min_{\theta}$$

Упр.

$$y = \langle w, x\rangle + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2) \quad \mu$$

$$y \mid x \sim \mathcal{N}(\underline{\langle w, x\rangle}, \sigma^2)$$

$$P(y \mid x, \theta) \qquad \theta = (w, \cancel{\sigma^2})$$

$$P(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

$$L(w) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(y_i - \langle w, x\rangle)^2}{2\sigma^2}\right) \longrightarrow \max_{\theta}$$

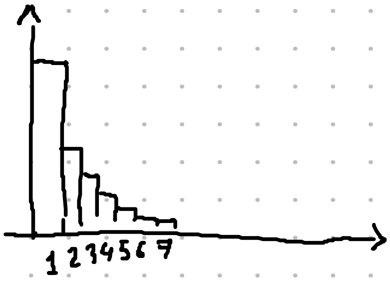
$$\ln L(w) = -n \cancel{\log\sqrt{2\pi\sigma^2}} - \cancel{\frac{1}{2\sigma^2}} \sum_{i=1}^{n} (y_i - \langle w, x\rangle)^2$$

$$\ln L(w) \propto \underbrace{-\frac{1}{n}\sum (y_i - \langle w, x\rangle)^2}_{-\,MSE}$$

Выводы:
- Нормальный шум соотвествует MSE, нормальное распределение не предполагает выбросов

- Лапласовский шум соотвествует MAE, он предполагает наличие выбросов (хвосты распределения толще) => модель с MAE может нормально работать с выбросами и не обращать на них внимание

Упр.



$$\mathbb{P}(y_i=k) = \frac{e^{-\lambda}\cdot\lambda^k}{k!}$$

$\lambda(x_i) =$ модель

$\langle w, x\rangle$

бустинг

$$L(w) = \mathbb{P}(y\,|\,x,w) = \prod_{i=1}^{n} \frac{e^{-\lambda(x_i)}\cdot(\lambda(x_i))^{y_i}}{y_i!}$$

$$(-1)\cdot\log L(w) \propto \sum\left(y_i\,\underline{\log\lambda(x_i)} - \lambda(x_i) - \cancel{\log y_i!}\right)$$

$$\text{poissloss} = \frac{1}{n}\sum\left[\lambda(x_i) - y_i\cdot\log\lambda(x_i)\right] \to \min_w$$

Когда мы доберемся до рекомендательных систем и будем говорить про ALS (матричные разложения), эта функция потерь ещё у нас всплывёт

Упр.

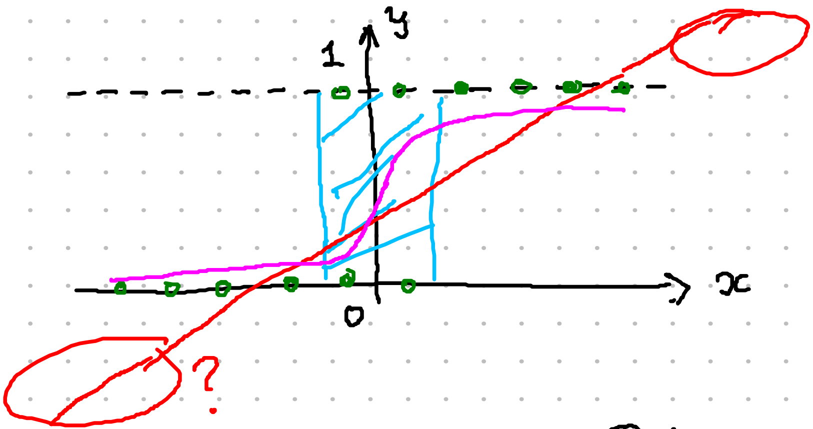
Y	0	1
P	$1-p$	p

$y_i = 1$ 10 раз

$y_i = 0$ 5 раз

$$L(p) = p^{10}(1-p)^{5}$$

$$p^{\sum y_i}(1-p)^{n-\sum y_i}$$



$$y = wx$$

$$F(wx)$$

$$F(t) = \frac{1}{1+e^{-t}}$$

$$P_i = \mathbb{P}(y_i = 1 \mid x) = F(\langle w, x_i \rangle + w_0)$$

y_i	0	1
	$1-P_i$	P_i

$$L = P_1^{y_1} \cdot (1-P_1)^{1-y_1} \cdot \ldots \cdot P_n^{y_n} \cdot (1-P_n)^{1-y_n}$$

$$\log L \propto \frac{1}{n} \sum \left[y_i \ln P_i + (1-y_i) \ln(1-P_i) \right]$$

$$\text{logloss} \propto -\log L \longrightarrow \min_w$$

$y_i \in \{0, 1\}$ $\quad y_i \log P_i + (1-y_i) \log(1-P_i)$

$y_i \in \{-1, 1\}$ $\quad \log\left[1 + \exp(y_i \langle w, x_i \rangle)\right]$

ММП обладает кучей хороших асимптотических свойств

Это работает если

$$\frac{n}{d} \longrightarrow \infty$$

$\boxed{\frac{n}{d} < 1}$ в жизни !!

$\Rightarrow$ Нужна регуляризация

2. Байесовский и частотный подходы

Частотный	Байесовский
θ − неизвестная константа	θ − случайная величина
Данные: $y_1, \dots y_n$	Данные: $y_1, \dots, y_n$
Модель: описывает связь параметров и данных	Ровно те же самые модели что и в частотном подходе

Частотный

ММП $\hat{\theta}^{ML}$

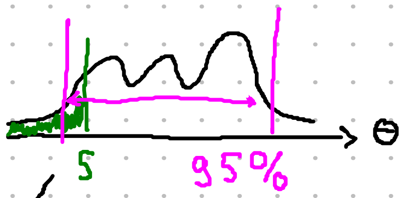
$\hat{\theta}^{ML} \overset{asy}{\sim} \mathcal{N}(\dots)$

$[\theta^L \ \theta^R]$ АБ-тест

$\hat{\theta} \xrightarrow{\text{модель}} \hat{y}_{n+1}$

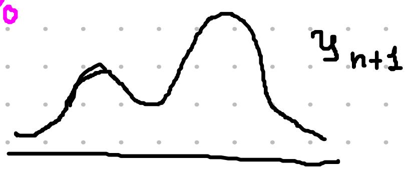
Байесовский

Байесовский вывод $p(\theta|y)$



$P(\theta < 5)$ −?

$\mathbb{E}(\theta|y)$



$p(y|x,\theta)$ — правдоподобие

$p(\theta)$ — априорное распределение

$$P(A|B) = \frac{P(AB)}{P(B)}$$

$p(\theta|x,y)$ — апостериорное распределение

$$p(\theta|x,y) = \frac{p(y,\theta|x)}{p(y|x)} = \frac{p(y|\theta,x)\cdot p(\theta)}{\underbrace{p(y|x)}_{\int p(y,\theta|x)\,d\theta}}$$

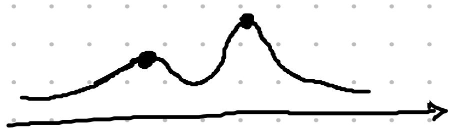
$$p(\theta)\cdot p(y|\theta,x)$$

не завистит от θ

$$p(\theta|x,y) \propto p(y|\theta,x)\cdot p(\theta)$$

Байес для бедных:
- апостерионое распределение искать дорого, зато легко найти его моду

$$\hat{\theta}^{MAP} = \mathrm{Mod}(\theta|x,y)$$



$$p(\theta|x,y) \to \max_{\theta}$$

$$\ln p(\theta|x,y) = \ln p(y|\theta,x) + \underline{\ln p(\theta)} \to \max_{\theta}$$

$$p(\theta) = 1 \quad \theta \sim U[-\infty;+\infty] \quad \hat{\theta}^{MAP} = \hat{\theta}^{ML}$$

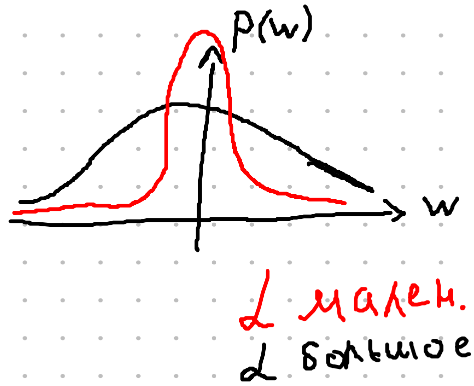
$$(-1)\cdot \text{posterior} = -\text{likelihood} + -\text{prior}$$

$$L(y,a) + R(\theta) \to \min_{\theta}$$

Упр.

$$y = \langle w, x \rangle + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$w_j \sim \mathcal{N}(0, \alpha^2)$$



$$P(w \mid y, x) \propto P(y \mid x, w) \cdot p(w) \longrightarrow \max_w$$

$$-\ln P(w \mid y, x) \propto \underbrace{\frac{1}{2\sigma^2}} \sum (y_i - \langle w, x \rangle)^2 + \overset{\cancel{2}\sigma^2}{\frac{1}{\cancel{2}\alpha^2}} \cdot \sum w_j^2 \longrightarrow \min_w$$

$$MSE + \frac{\sigma^2}{\alpha^2} \cdot \|w\|^2 \qquad \text{Ridge}$$

$$\sigma^2 = 1 \qquad 1/\alpha^2 = \lambda$$

Что надо для LASSO?

Упр.

y_1 – только одно набл.

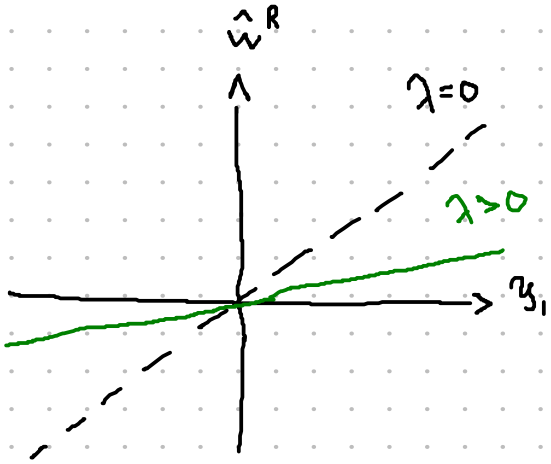
$y = w$

Ridge

$Q = (y_1 - w)^2 + \lambda \cdot w^2 \to \min\limits_w$

$-2(y_1 - w) + 2\lambda w = 0$

$\hat{w}^R = \dfrac{y_1}{1+\lambda}$



LASSO

$Q = (y_1 - w)^2 + \lambda \cdot |w| \to \min\limits_w$

$-2(y_1 - w) + \lambda \cdot \operatorname{sgn} w$

- $w > 0$ $\quad w^+ = y_1 - \frac{\lambda}{2} \Leftrightarrow y_1 > \frac{\lambda}{2}$
- $w < 0$ $\quad w^- = y_1 + \frac{\lambda}{2} \Leftrightarrow y_1 < -\frac{\lambda}{2}$

$y_1 + \frac{\lambda}{2} < 0$

$w^- > w^+ \quad \lambda \neq 0$

$y_1 \in [-\frac{\lambda}{2}; \frac{\lambda}{2}]$??? $w = 0$

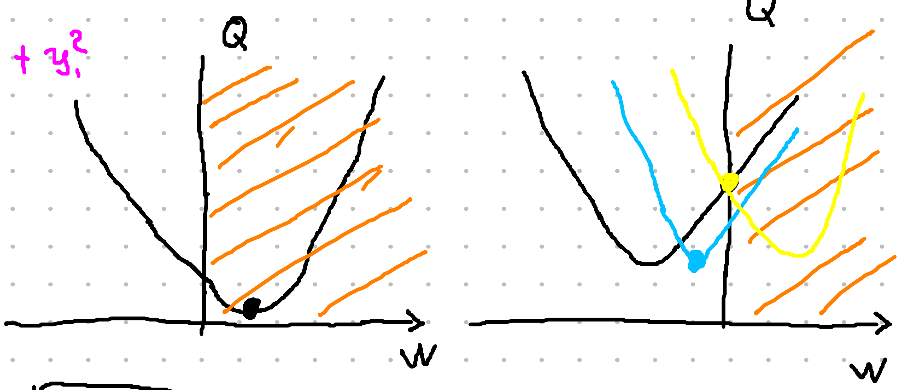
$Q = y_1^2 - 2y_1 w + w^2 \pm \lambda \cdot \cancel{\ldots} \cdot w$

$w^2 - (2y_1 \mp \lambda) w + y_1^2$

$aw^2 + bw + c$

парабола

$w^+ > 0$ $\varnothing$ $w^- > 0$	$w^+ < 0$ ② $w^- < 0$ $\varnothing$
$w^+ < 0$ $w^- > 0$ $\quad 0$	$w^+ > 0$ $w^- < 0$ $\quad \varnothing$



$\boxed{w > 0}$

$$\hat{w} = \begin{cases} y_1 - \frac{\lambda}{2}, & y_1 > \frac{\lambda}{2} \\ y_1 + \frac{\lambda}{2}, & y_1 < -\frac{\lambda}{2} \\ 0 & \text{иначе} \end{cases}$$