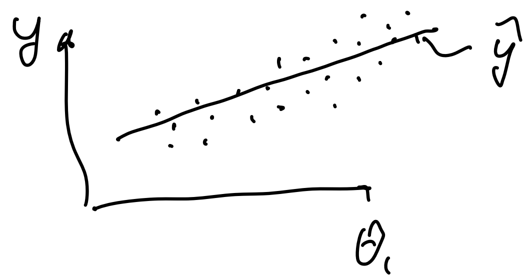
$$y = \underbrace{X_{n\times m}\,\theta_{m\times 1}}_{\hat{y}} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$



$$\boxed{\theta - ?}$$

$$MSE = \sum_{i=1}^{n} (y_i - x_i\theta)^2 \to \min_{\theta}$$

$$\begin{cases} MSE \to \min_{\theta} \\ \|\theta\|_{1,2} < C \end{cases}$$

L_1 – Lasso – Лассо

L_2 – Ridge – Гребневая рег

1. Lasso / Ridge

Почему спарсификации при Lasso??

$\theta = ((1, \varepsilon))$ $\quad \varepsilon \ll 1 \quad \|\theta\| = \|(1, \varepsilon)\| < C$

$\to (1, 0)$

$\boxed{1.1}$

$$\|\underset{=(1,\varepsilon)}{\theta} - (\delta, 0)\|_2^2 = \boxed{1 - 2\delta + \delta^2 + \varepsilon^2}$$

$$\|\theta - (\delta, 0)\|_1 = \boxed{1 - \delta + \varepsilon}$$

$\boxed{1.2.}$

$$\|\theta - (0,\delta)\|_2^2 = \boxed{1 - 2\varepsilon\delta + \delta^2 + \varepsilon^2}$$

$$\|\theta - (0,\delta)\|_1 = \boxed{1 - \delta + \varepsilon}$$

$$2\delta \lor 2\varepsilon\delta$$

$$\underline{2\delta} > 2\varepsilon\delta$$

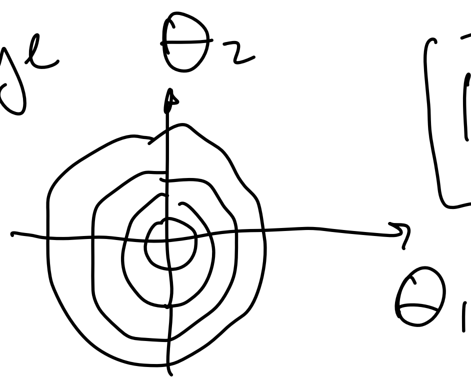
2. $\begin{cases} MSE \to \min\limits_{\theta} \\ \boxed{\|\theta\|_{1,2} < C} \end{cases}$

$\theta = (\theta_1, \theta_2)$

$\theta = \theta_1$

$MSE = (y - X\theta_1)^2$

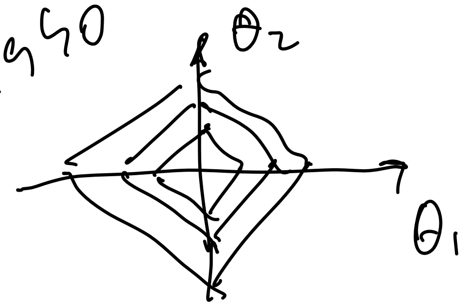
Ridge



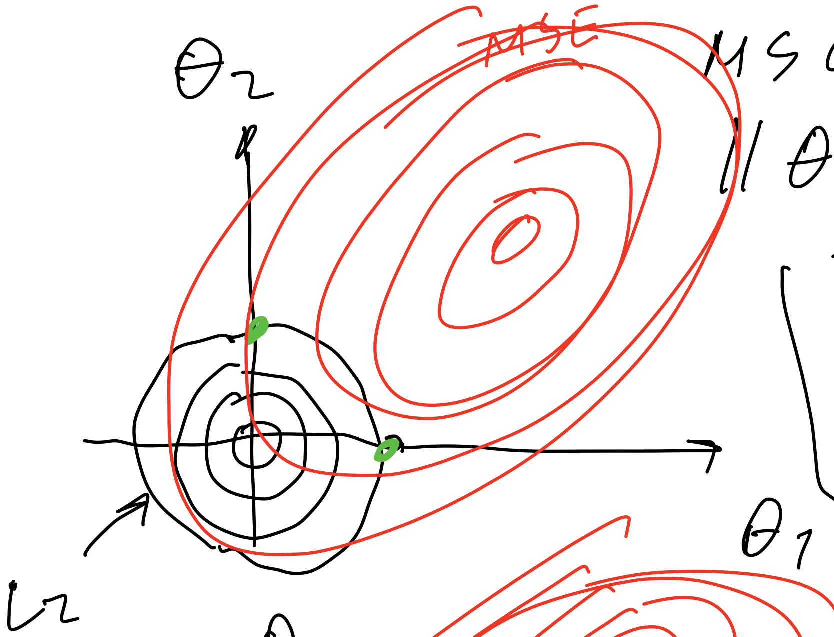
$\boxed{\|\theta\|_2^2} < C$

$a^2 + b^2 = C$
$= r$

Lasso



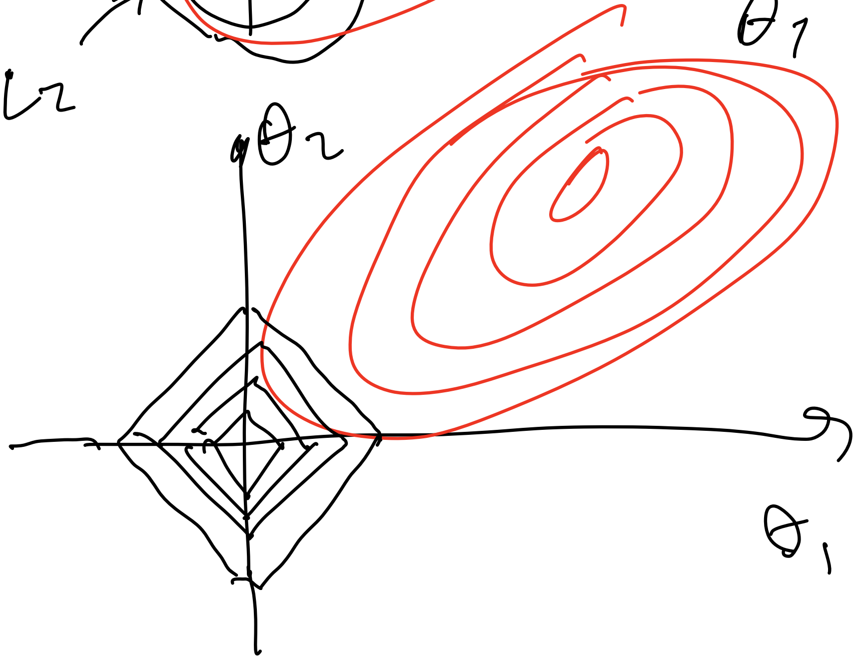
$\|\theta\|_1 < C$



$MSE \to \min$

$\|\theta\|_2^2 < C$

Ridge



Lasso

y

Средняя продажи. низкие

↑

Почтовый индекс

$$\begin{cases} MSE \to \min\limits_{\theta} \\ \|\theta\|_2^2 < C_1 \\ \|\theta\|_1 < C_2 \end{cases}$$

Elastic Net.

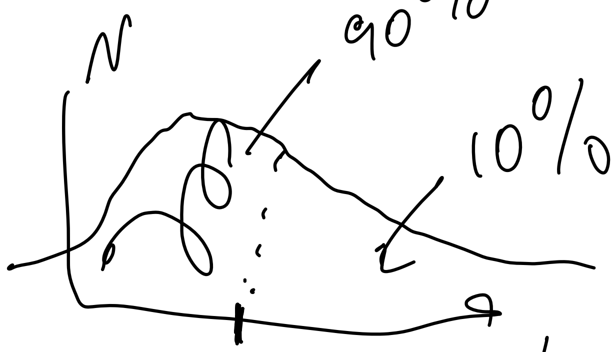
MLE, MAP

$P(y|\theta)$ — ? Частотники

$$\boxed{p(\theta|y) = \frac{p(y|\theta)\cdot p(\theta)}{p(y)}}$$

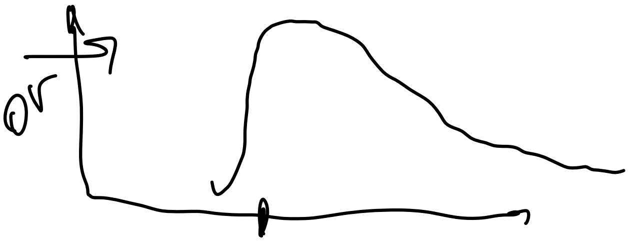
$$\text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidance}}$$

prior →



y

posterior →



$$p(y|\theta) \rightarrow \text{max likelihood}$$

$$\underset{\theta}{\operatorname{argmax}} \boxed{P(y|\theta, \sigma^2)} =$$

$$\boxed{y = \underline{X\theta} + \mathcal{N}(0, \sigma^2)}$$

$$= \prod_{i=1}^{n} \boxed{P_y(y_i | \theta, \sigma^2)} = \mathcal{N}(X\theta, \sigma^2)$$

$$=$$

$$\mathcal{N}(0, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}$$

1

$$= \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i - x_i\theta)^2/2\sigma^2} \right) \rightarrow \max_{\theta}$$

$a \cdot b$

$\log a \cdot b =$

$\log a + \log b$

likelihood

log likelihood

$\log \prod (\ldots) =$

$= \sum \log (\ldots) =$

$$= \boxed{\sum \log \left(e^{-(y_i - x_i \theta)^2 / 2\sigma^2} \right) + \cancel{\sum_{i=1}^{n} \log \left(\frac{1}{\sigma \sqrt{2\pi}} \right)}}$$

$\max_{\theta}$

$$= \boxed{- \sum (y_i - x_i \theta)^2} + \cancel{\sum 2\sigma^2}$$

$\max_{\theta}$

$$\boxed{\sum_{i=1}^{n} (y_i - x_i\theta)^2 \to \min_{\theta}}$$

$$y = X\theta + \varepsilon$$

MSE

MLE

L_1, L_2

MAP

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

$$p(\theta|y) = \frac{p(y|\theta)\cdot p(\theta)}{p(y)} \to \max_{\theta}$$

$$\Updownarrow$$

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

$$\underbrace{\log P(y \mid \theta)}_{\text{likelihood}} + \underbrace{\log P(\theta)}_{\text{prior}} \to \max_{\theta}$$

L_2 , L_1

$\theta \sim \mathcal{N}(0, \hat{\sigma}^2)$ $\theta \sim$ Laplace

$$MSE + \lambda \sum \theta_i^2 = MAP\ L_2$$

$$MSE + \lambda \sum |\theta_i| = MAP\ L_1$$

$$p(y \mid X) = \frac{p(y) \cdot p(X \mid y)}{p(X)} \sim$$

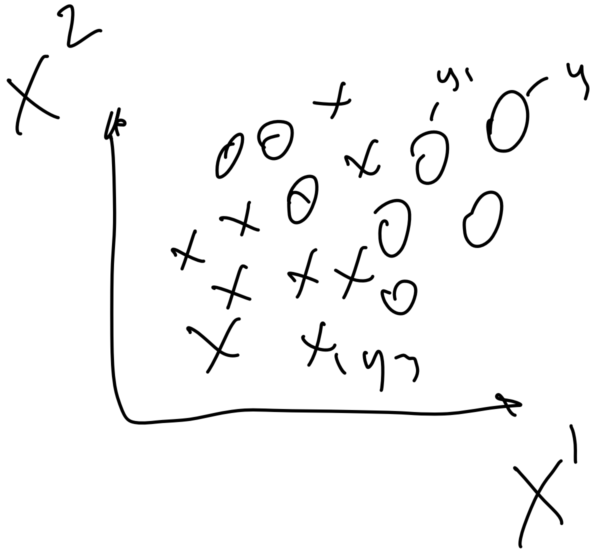
$$\sim p(y) \cdot p(x \mid y) = p(y)\, p(x^1, x^2, \ldots x^n \mid y)$$

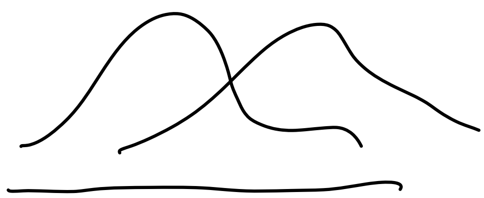
$$= p(y) \cdot p(x^1 | y) \cdot p(x^2, \dots, x^m \mid y, x^1) =$$

$$= p(y) \cdot p(x^1 | y) \cdot p(x^2 \mid y, x^1) \cdot p(x^3, \dots, x^m \mid y, x^1, x^2) =$$

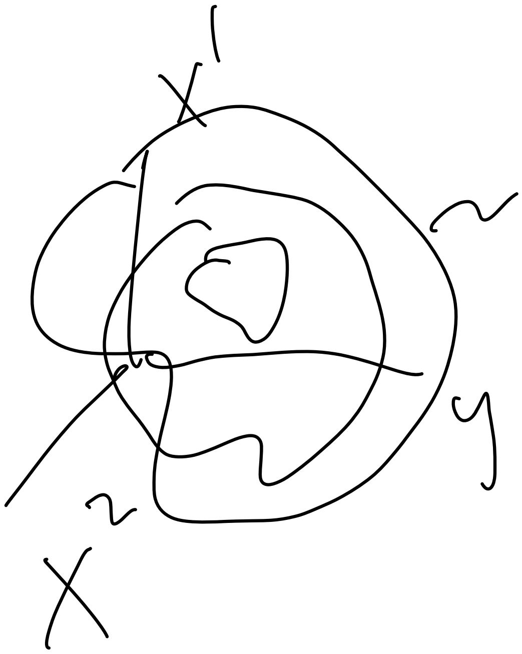
$$= \dots \dots = p(y) \cdot p(x^1 | y) \cdot p(x^2 \mid y, x^1) \cdot p(x^3 \mid y, x^1, x^2) \cdot \dots \cdot p(x^m \mid y, x^1, x^2, \dots, x^{m-1}) =$$

$y = 0 : x^1 ?$





$P(x^2 \mid y, x^1)$



$$p(x^3 \mid x^1, x^2, y) =$$

$$= p(x^3 \mid y)$$

$$= \quad \approx \quad p(y) \cdot p(x^1 \mid y) \cdot p(x^2 \mid y) \cdot p(x^3 \mid y) \cdot \ldots \cdot p(x^n \mid y)$$