Statistical decision theory

Регрессия: $L(y, \hat{y}) = \underbrace{(y - \hat{y})^2}_{\to \min}$

правд. прогноз.

$\bar{x} \xrightarrow{f} \hat{y}$ $\quad D \sim p(\bar{x}, y)$

$\quad\quad " \{(\bar{x}_n, y_n)\}_{n=1}^N$

$$EPE[f] = \mathbb{E}_{p(\bar{x}, y)}\left[(y - f(\bar{x}))^2\right] \to \min$$

$$\| \quad \int (y - f(\bar{x}))^2 \, p(\bar{x}, y) \, d\bar{x} \, dy =$$

$$= \int (y - f(\bar{x}))^2 \, p(y|\bar{x}) \, p(\bar{x}) \, d\bar{x} \, dy =$$

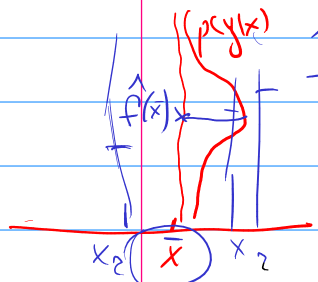
$$= \int \left[\int (y - f(\bar{x}))^2 p(y|\bar{x}) \, dy \right] p(\bar{x}) \, d\bar{x}$$

$f(\bar{x}) \to \min$

$$\int (y - f(\bar{x}))^2 p(y|\bar{x}) \, dy \xrightarrow[f(\bar{x})]{} \min$$

$$\mathbb{E}_{p(y|\bar{x})}\left[(y - a)^2\right] \xrightarrow[a]{} \min$$

$$\boxed{\hat{f}(\bar{x}) = \mathbb{E}_{p(y|\bar{x})}[y]}$$ функция регрессии



$$\hat{f}(\bar{x}) = \mathbb{E}_{p(y|\bar{x})}[y] \approx$$

$$\approx \frac{1}{R} \sum_{r=1}^{R} y^{(r)} \quad y^{(r)} \sim p(y|\bar{x})$$

$$\approx \frac{1}{R} \sum y_r \quad , y_r \sim p(y|\bar{x}_r)$$

$$\bar{x}_r \in kNN(\bar{x})$$

$y \in 1 \dots K$

Классификация:

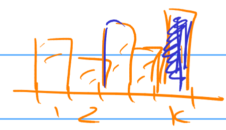
$$L(y, \hat{y}) = \begin{cases} 1, & y \neq \hat{y} \\ 0 & y = \hat{y} \end{cases} \to \min$$

$$EPE[f] = \mathbb{E}_{p(\bar{x}, y)}\left[L(y, f(\bar{x}))\right] \xrightarrow[f]{} \min$$

$$\| \quad \int \left[\sum_{k=1}^{K} L(y = k, f(\bar{x})) \, p(y = k|\bar{x}) \right] p(\bar{x}) \, d\bar{x}$$

$$\xrightarrow[f(\bar{x})]{} \min$$

$$\sum_{k=1}^{K} [f(\bar{x}) \neq k] \, p(y = k|\bar{x}) \xrightarrow[f(\bar{x})]{} \min$$



$$\boxed{\hat{f}(\bar{x}) = \arg\max_k p(y = k|\bar{x})}$$

$$L(y, \hat{y}) = \begin{array}{c|cc} & y & \hat{y} \\ \text{рак} & 0 & 1000 \\ \text{нет} & 1 & 0 \end{array}$$

$$\sum_{k=1}^{K} L(y = k, f(\bar{x})) \, p(y = k|\bar{x}) \to \min$$

$$\hat{f} = \arg\max_k \left[p(y = k|\bar{x}) \cdot L(\ldots)\right]$$

$f \to \min$ $\quad +\hat{f}(\bar{x}) - \hat{f}(\bar{x})$ $= \left((y - \hat{f}(\bar{x})) + (\hat{f}(\bar{x}) - f(\bar{x}))\right)^2$

$$EPE[f] = \int (y - f(\bar{x}))^2 \, p(\bar{x}, y) \, d\bar{x} \, dy =$$

$$= \int (y - \hat{f}(\bar{x}))^2 p(\bar{x}, y) \, d\bar{x} \, dy + 2 \int (y - \hat{f}(\bar{x}))(\hat{f}(\bar{x}) - f(\bar{x})) \, p(\bar{x}, y) \, d\bar{x} \, dy + \int (\hat{f}(\bar{x}) - f(\bar{x}))^2 p(\bar{x}, y) \, d\bar{x} \, dy$$

$p(y|\bar{x}) p(\bar{x})$

$$2 \int \left(\int (y - \hat{f}(\bar{x})) \, p(y|\bar{x}) \, dy \right) (\hat{f} - f) \, p(\bar{x}) \, d\bar{x}$$

$$\mathbb{E}_{p(y|\bar{x})} y$$

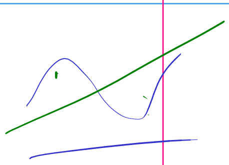
$$\int y \, p(y|\bar{x}) \, dy - \mathbb{E} y = 0$$

NOISE

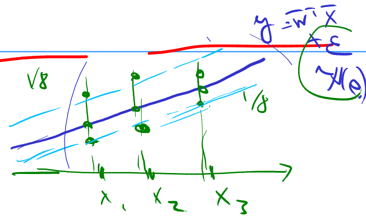
$$= \boxed{\int (y - \hat{f}(\bar{x}))^2 p(\bar{x}, y) \, d\bar{x} \, dy}$$

$$+ \int (\hat{f}(\bar{x}) - f(\bar{x}))^2 p(\bar{x}, y) \, d\bar{x} \, dy$$

$f \to \min$



$$\boxed{f(\bar{x}; D)} \quad D \sim p(\bar{x}, y)$$



$$\int (\hat{f} - f)^2 p(\bar{x}, y)\, d\bar{x}\, dy = \int (\hat{f} - \mathbb{E}f)^2 p(\bar{x}, y)\, d\bar{x}\, dy + 2\int (\hat{f} - \mathbb{E}f)(\mathbb{E}f - f)\, p(\bar{x}, y)\, d\bar{x}\, dy + \int (f - \mathbb{E}f)^2 p(\bar{x}, y)\, d\bar{x}\, dy$$

$$+ \mathbb{E}_{D \sim p(\bar{x}, y)} f(\bar{x}; D) - \mathbb{E}_D f$$

$$\int_D (\mathbb{E}_D f - f)\, p(\bar{x}, y)\, d\bar{x}\, dy = 0$$

$$EPE[f] = \underbrace{\int (\mathbb{E}_D f(\bar{x}; D) - \hat{f}(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy}_{\text{Bias}} + \underbrace{\int (f(\bar{x}; D) - \mathbb{E}_D f(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy}_{\text{Variance}} + \underbrace{\int (y - \hat{f}(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy}_{\text{Noise}}$$

Bias — Variance — Noise

decomposition

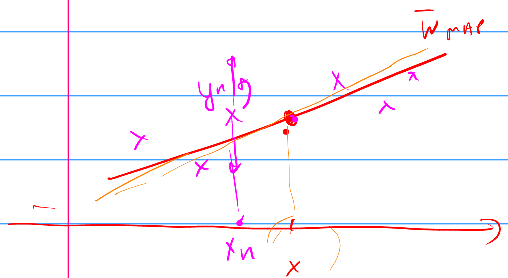
$$p(y \mid \bar{x}, D) = \int p(y \mid \bar{w}, \bar{x})\, p(\bar{w} \mid D)\, d\bar{w} = \mathcal{N}(y \mid \bar{w}_{MAP}^T \bar{x}, \sigma^2 + \dots)$$

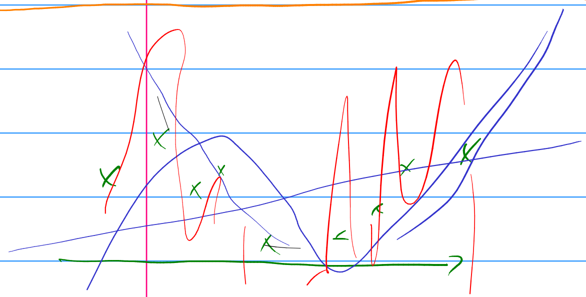
$$\bar{w}_{MAP} = \frac{1}{\sigma^2} \Sigma_N X^T \bar{y}$$

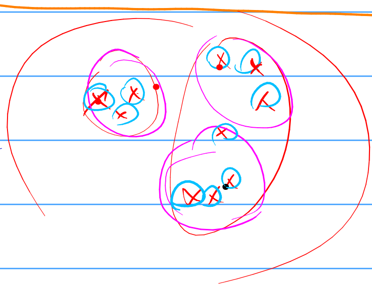
$$y_{MAP}(\bar{x}) = \frac{1}{\sigma^2} (\Sigma_N X^T \bar{y})^T \bar{x} = \left(\frac{1}{\sigma^2} \cdot \bar{x}^T \Sigma_N X^T\right) \bar{y} = \bar{a}^T \bar{y}$$

$$= \sum_{n=1}^{N} \left(\underbrace{\frac{1}{\sigma^2} \bar{x}^T \Sigma_N \bar{x}_n}_{k(\bar{x}, \bar{x}_n)} \right) y_n = \sum_{n=1}^{N} k(\bar{x}, \bar{x}_n)\, y_n$$

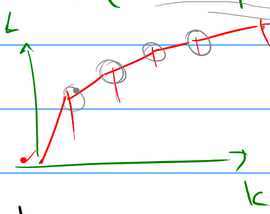
equivalent kernel







$$L(\bar{w}) = p(D \mid \bar{w})$$



Model selection

$D \quad \mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_k$

$$p(\mathcal{M}_i \mid D) \propto p(\mathcal{M}_i)\, p(D \mid \mathcal{M}_i)$$

$$p(\theta \mid D)_{\mathcal{M}} = \frac{p(\theta)\, p(D \mid \theta)_{\mathcal{M}}}{p(D)_{\mathcal{M}}}$$

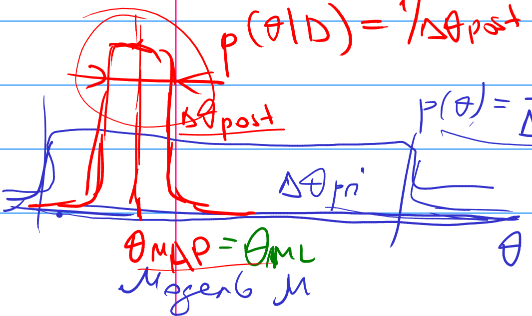
$$p(D \mid \mathcal{M}_i) = \int p(D \mid \theta, \mathcal{M}_i)\, p(\theta \mid \mathcal{M}_i)\, d\theta =$$

$$p(\theta \mid D) = 1/\Delta\theta_{post} \propto p(\theta) p(D \mid \theta)$$

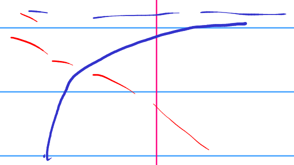
$$= \int_{prior} \frac{1}{\Delta\theta_{pri}} \cdot p(D \mid \theta, \mathcal{M}_i)\, d\theta =$$

$$= \frac{1}{\Delta\theta_{pri}} \int_{post} p(D \mid \theta_{ML}, \mathcal{M}_i)\, d\theta =$$

$$= p(D \mid \theta_{ML}, \mathcal{M}_i) \cdot \left(\frac{\Delta\theta_{post}}{\Delta\theta_{pri}}\right)^{M}$$



$$\log p(M_i|D) = \text{Const} + \log p(M_i) + \log \frac{\Delta\theta_{post}(M_i)}{\Delta\theta_{pri}(M_i)} + \log p(D|\theta_{ML}, M_i) =$$



$$= \log p(D|\theta_{ML}, M_i) - \left(M \log \frac{\Delta\theta_{pri}}{\Delta\theta_{post}} - \log p(M_i) \right) + \text{const}$$

BIC

$$BIC(M) = \log p(D|\theta_{ML}, M) - \frac{1}{2}|M| \ln N$$

$$= \sum \log p(d|\ )$$

$$p(D|M) \xrightarrow[M]{} \max \qquad D \sim p(x|M^*)$$

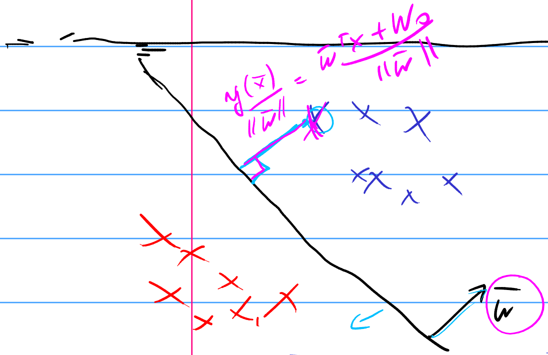
$$\mathbb{E}_{D\sim M^*}[\log p(D|M)] = \int \log \frac{p(D|M)}{p(D|M^*)}\, p(D|M^*)\, dD = -KL\big(p(D|M^*)\,\|\,p(D|M)\big) \le 0$$

$$\pm \int \log p(D|M^*)\, p(D|M^*)\, dD$$

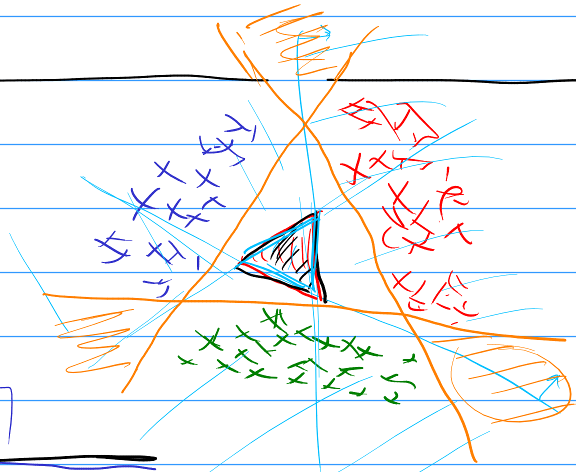
$$KL(p\|q) = \int \log \frac{p}{q}\, p\, dx$$

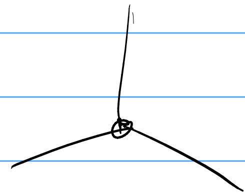
≥ 0

$= 0$ iff $p = q$



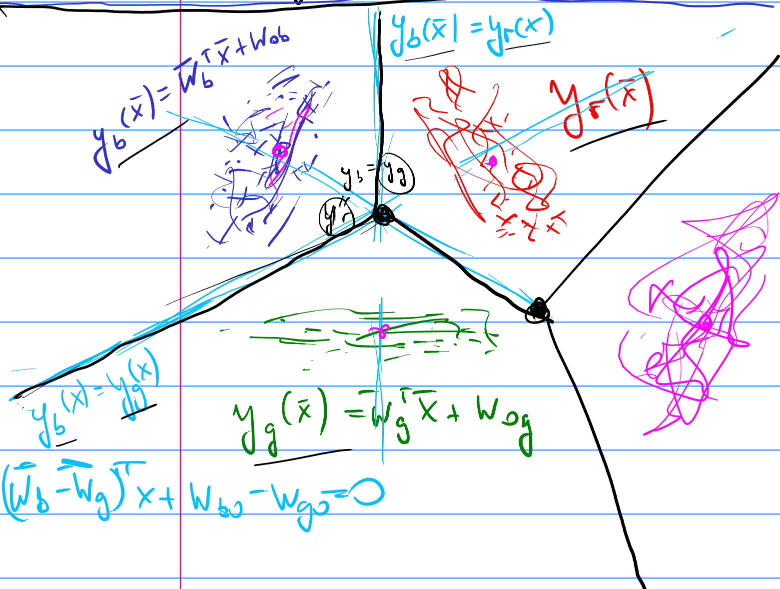
$$y(\bar x) = \bar w^T \bar x + w_0 = 0$$





$\max\{y_1, y_2, y_3\}$

где единств.



$$(\bar w_b - \bar w_g)^T x + w_{b0} - w_{g0} = 0$$