1) Фреквентистский подход.

n - частота

MLE $\quad P(X \mid \Theta)$ - likelihood

Пример.

$$P(X) = \mu^{x} \cdot (1-\mu)^{(1-x)}$$

1 $\quad$ 0

3 $\quad$ 5. $\qquad$ iid.

$$\text{MLE}: \quad \Theta = \arg\max_{\Theta} \sum_{x} \Theta^{x} \cdot (1-\Theta)^{(1-x)}$$

$$\log\left(\Theta^{x} \cdot (1-\Theta)^{(1-x)}\right) = x \log \Theta +$$

$$+ (1-x) \log(1-\Theta) = \sum_{x} \Big(x \log \Theta + \log(1-\Theta)$$

$$- x \log(1-\Theta)\Big) =$$

$$f'(\theta) = \sum_x x \frac{1}{\theta} - \frac{n}{1-\theta} + \sum_x x \frac{1}{1-\theta} =$$

$$= \sum x \frac{1}{\theta - \theta^2} - \frac{n\theta}{\theta - \theta^2} = \frac{\sum_n x - n\theta}{\theta - \theta^2}$$

$$n\theta = \sum_n x \qquad \boxed{\theta = \frac{1}{n}\sum x}$$

---

| 1 | ∞ |
|---|---|
| 5 | 0 |

$\theta = 1$ $\boxed{\text{yne}}$

---

# BAYESIAN

likelihood

Prior

$$P(\theta \mid X) = \frac{P(X \mid \theta) \cdot P(\theta)}{P(X)}$$

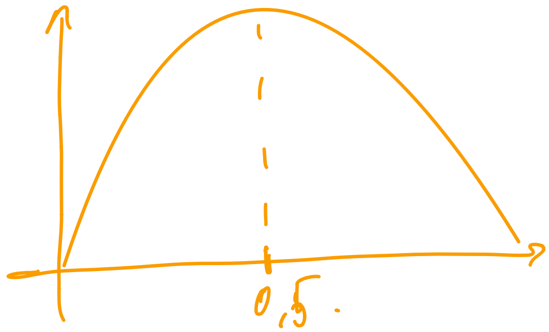
evidence

aposteriori

beta distribution

$$P(x|\alpha,\beta) = \overset{= \text{constant}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}} x^{\alpha-1}(1-x)^{\beta-1}$$

$\alpha = \beta = 3.$




$P(\theta) = \theta^2(1-\theta)^2.$

$$\log P(\theta|X) = \sum(x\log\theta + (1-x)\log(1-\theta))$$

$$+2\log \overset{\theta-\theta^2}{\theta(1-\theta)}.$$

$$f'(\theta) = \sum_x \left( x\frac{1}{\theta} - \frac{(1-x)}{1-\theta} \right) + 2\frac{1}{\theta-\theta^2}\cdot(1-2\theta)$$

$$\sum_x \frac{x(1-\theta) - (1-x)\theta}{\theta-\theta^2} = \frac{x - \cancel{x\theta} - \theta + \cancel{x\theta}}{\theta-\theta^2} +$$

$$+\frac{2-4\theta}{\theta-\theta^2} = \sum_x x - [n\theta - 2 + 4\theta]$$

$$n\theta - 2 + 4\theta = \sum x$$

$$(n+4)\theta = \sum x + 2$$

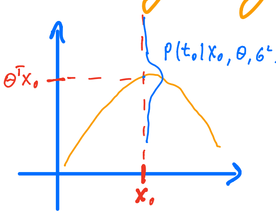
$$\theta = \frac{(\sum x)+2}{n+4}. \qquad \frac{5+2}{5+4} = \boxed{\frac{7}{9}}$$

$\boxed{\text{data overwhelm prior}}$

So how about some regression?

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

для данной точки $x$ целевая переменная $f(x)$ распределена по Гауссу со средним $\theta^T x$.




или что эквивалентно

$$f(x) = \theta^T x + \xi, \text{ где } \xi \sim N(0, \sigma^2)$$

$$p(t \mid x, \theta, \sigma^2) = N(t \mid \theta^T x, \sigma^2) =$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-\theta^T x)^2}.$$

$$\hat{\theta} = \arg\max_{\theta} \ln \left( \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t_i - \theta^T x_i)^2} \right) =$$

$$= -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} (t_i - \theta^T x_i)^2\right] - \frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi.$$

It's equivalent to MSE minimization

$$\sum_{i=1}^{n} (t_i - \theta^T x_i)^2.$$

Знаем как можем найти $\sigma^2$.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (t_i - \hat{\theta}^T x_i)^2$$

Predictive distribution.

Bayesian regression

$$P(\theta) = N(\theta \mid 0, \alpha^{-1} I) \qquad P(\theta) = N(\theta \mid \mu_0, \sigma_0)$$

$$P(t \mid \theta^T x, \sigma^2) = \prod_{i=1}^{n} N(t_i \mid \theta^T x_i, \sigma^2)$$

$$P(\theta|t) = \frac{P(t|\theta' x, \sigma^2)\, P(\theta)}{P(t)}.$$

$$\ln P(\theta|t) = \frac{1}{2\sigma^2} \sum_{i=1}^{n} (t_i - \theta^T x_i)^2 - w^T w \cdot \frac{\alpha}{2} + const.$$

---

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

$$N(t|\theta^T x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(t-\theta^T x)^2}$$

$$N(\theta|0, \underset{(m\times m)}{I/\alpha}) = \frac{1}{\sqrt{(2\pi)^m |I| \cdot \frac{1}{\alpha}}} \cdot e^{-\frac{1}{2}(\theta-0)^T I^{-1} (\theta-0)\,\alpha}$$

$$= \frac{\sqrt{\alpha}}{\sqrt{(2\pi)^m |I|}} \cdot e^{-\frac{\alpha}{2}\theta^T \cdot I^{-1=I} \theta} = \frac{\sqrt{\alpha}}{\sqrt{(2\pi)^m}} \cdot e^{-\frac{\alpha}{2}\theta^T \cdot \theta}$$

$$P(\theta|t, \theta^T x, \sigma^2) = \frac{\sqrt{\alpha}}{\sqrt{(2\pi)^{m+1}} \cdot \sqrt{\sigma^2}} \boxed{\cdot e^{-\frac{1}{2\sigma^2}(t-\theta^T x)^2 - \frac{\alpha}{2}\theta^T\theta}}$$

---

ln

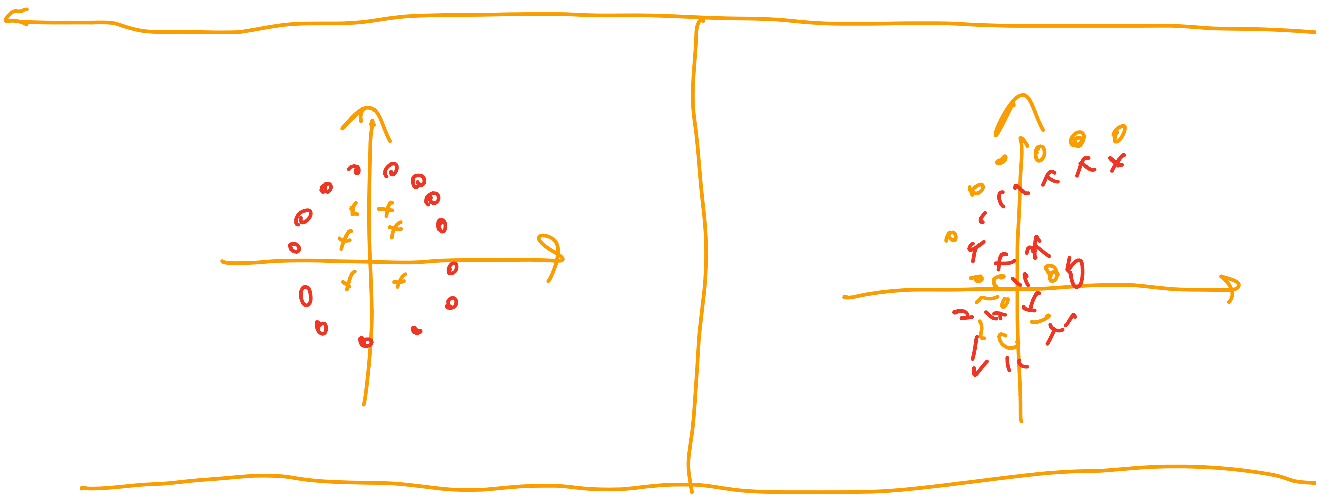
exponential family:

basis function

$$t = \psi(\theta^T \varphi(x))$$

activation

$\psi^{-1}$ — lina. function.



$$\ldots \psi_2(\theta_2^T \varphi_2(\psi_1(\theta_1^T \varphi_1(x))))$$

1) Local minima.
2) Saddle points.
3).

$$p(x|\eta) = h(x)g(\eta)e^{\eta^T u(x)}$$

$g(\eta)$ - norm coefficient

$\eta$ - natural parameters

$u(x)$

$$p(x|\mu) = \mu^x(1-\mu)^{(1-x)}$$

$$p(x|\mu) = e^{x\ln\mu + (1-x)\ln(1-\mu)} =$$

$$= e^{x\ln\mu + \ln(1-\mu) - x\ln(1-\mu)} =$$

$$= (1-\mu)e^{\ln\left(\frac{\mu}{1-\mu}\right)x}$$

$$\eta = \ln\left(\frac{\mu}{1-\mu}\right) \quad e^{\eta} = \frac{\mu}{1-\mu}$$

$$\mu = \frac{1}{1+\exp(-\eta)} = \sigma(\eta).$$

$$\boxed{P(x|\eta) = \sigma(-\eta)e^{\eta x}.}$$

$u(x) = x \qquad g(\eta) = \sigma(-\eta).$

$h(x) = 1$