

17-11-03.2022 | neural networks

$$y_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = f(x_i) \quad , \quad \text{data } y_1, \dots, y_n, x_1, \dots, x_n$$

$$P(y_1, \dots, y_n | x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(f(x_i) - y_i)^2}{2\sigma^2}}$$

$$\log P(y_1, \dots, y_n | x_1, \dots, x_n) = \sum_{i=1}^n -\log(\sqrt{2\pi}\sigma) -$$

$$- \frac{(f(x_i) - y_i)^2}{2\sigma^2} \quad \text{assume } \sigma \text{ is known}$$

$$\arg \max_{\vec{\beta}, \vec{\gamma}} \log P(y_1, \dots, y_n | x_1, \dots, x_n)$$

$$= \arg \max_{\vec{\beta}, \vec{\gamma}} \left(\sum_{i=1}^n -\log(\sqrt{2\pi}\sigma) - \right.$$

$$\left. - \frac{(f(x_i) - y_i)^2}{2\sigma^2} \right) =$$

$$= \arg \max_{\vec{\beta}, \vec{\gamma}} - \sum_{i=1}^n (f(x_i) - y_i)^2$$

$$\vec{y}_1, \dots, \vec{y}_n, x_1, \dots, x_n$$

$$\vec{y}_i \sim N_k(\vec{\mu}_i = \vec{f}(x_i), \Sigma)$$

$$P(\vec{y}_1, \dots, \vec{y}_n | x_1, \dots, x_n) =$$

$$= \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{f}(x_i) - \vec{y}_i)^T \Sigma^{-1} (\vec{f}(x_i) - \vec{y}_i)}$$

$$\log P(\vec{y}_1, \dots, \vec{y}_n | x_1, \dots, x_n) =$$

$$= C + \sum_{i=1}^n -\frac{1}{2} (\vec{f}(x_i) - \vec{y}_i)^T \Sigma^{-1} (\vec{f}(x_i) - \vec{y}_i)$$

assume Σ is known

$$\vec{\gamma}, \vec{\beta} \rightarrow \max$$

$$\vec{y}_1, \dots, \vec{y}_n | x_1, \dots, x_n =$$

$$= (\vec{y}_{g1}, \vec{y}_{m1}), \dots, (\vec{y}_{gn}, \vec{y}_{mn})$$

continuous matrix class $x_1, \dots, x_n$ further assume

$$P(\vec{y}_{gi}, \vec{y}_{mi} | x_i) = P(\vec{y}_{gi} | x_i) P(\vec{y}_{mi} | x_i)$$

$$P(\vec{y}_1, \dots, \vec{y}_n | x_1, \dots, x_n) =$$

$$= \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{f}(x_i) - \vec{y}_i)^T \Sigma^{-1} (\vec{f}(x_i) - \vec{y}_i)}$$

$$\times C_i \prod_{j=1}^K P_{ij}^{\vec{y}_{mij}} \xrightarrow{\vec{\gamma}, \vec{\beta}} \max$$

$$P_{ij} = f_{mij}(x_i)$$

$y_i \overset{i.i.d.}{\sim} \text{Bernoulli}(p_i)$

$p_i = f(x_i)$. for the data $y_1, \dots, y_n, x_1, \dots, x_n$

$$P(y_1, \dots, y_n | x_1, \dots, x_n) = \prod_{i=1}^n f(x_i)^{y_i} (1 - f(x_i))^{1-y_i}$$

$$\log P(y_1, \dots, y_n | x_1, \dots, x_n) =$$

$$= \sum_{i=1}^n y_i \log f(x_i) + (1 - y_i) \log (1 - f(x_i))$$

$\vec{y}_1, \dots, \vec{y}_n, x_1, \dots, x_n$

$$P(\vec{y}_1, \dots, \vec{y}_n | x_1, \dots, x_n) =$$

$$= \prod_{i=1}^n C_i \prod_{j=1}^K p_{ij}^{y_{ij}} \quad (\Leftarrow)$$

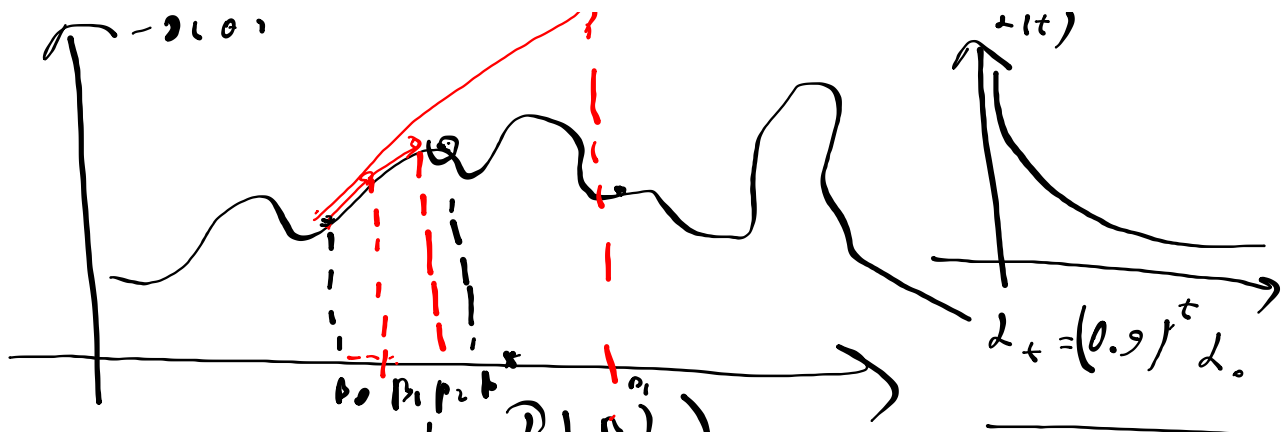
$$\log P(\vec{y}_1, \dots, \vec{y}_n | x_1, \dots, x_n) =$$

$$= \sum_{i=1}^n \log C_i + \sum_{j=1}^K y_{ij} \log p_{ij} \quad (\Leftarrow)$$

$$(\Leftarrow) \arg \max_{\vec{\beta}, \vec{\alpha}} \log P(\vec{y}_1, \dots, \vec{y}_n | x_1, \dots, x_n)$$

$$= \arg \max_{\vec{\beta}, \vec{\alpha}} \sum_{i=1}^n \log C_i + \sum_{j=1}^K y_{ij} \log f_j(x_i)$$

$$= \arg \max_{\vec{\beta}, \vec{\alpha}} \sum_{i=1}^n \sum_{j=1}^K y_{ij} \log f_j(x_i)$$



$$\beta_1 = \beta_0 + \lambda_1 \left(-\frac{\partial l(\beta)}{\partial \beta} \right)$$

$$\beta_2 = \beta_1 + \lambda_2 \left(-\frac{\partial l(\beta)}{\partial \beta} \right)$$

$$l(\beta) = -\mathcal{D}(\beta)$$

$$\hat{l}(\beta) = \frac{l^{(0)}(\beta)}{0!} + \frac{l'(\beta)\Delta}{1!} + \frac{l''(\beta)\Delta^2}{2!} =$$

$$\frac{\partial \hat{l}}{\partial \Delta}(\beta) = l'(\beta) + \Delta l''(\beta) (=)$$

$$(=) \Rightarrow \Delta = -\frac{l'(\beta)}{l''(\beta)} = -\frac{\mathcal{D}'(\beta)}{\mathcal{D}''(\beta)}$$