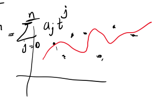
Ausgewählte Kapitel der Parameterschätzung
09.06.2021

posterior $\leftarrow p(x|y) = \dfrac{\overset{\text{likelihood function}}{p(y|x)}\, p(x)}{p(y)}$ → prior

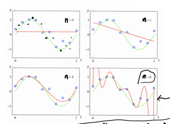
$y = a_0 + a_1 t_1 + a_2 t_2^2 + \dots + a_n t^n = \sum_{j=0}^{n} a_j t^j$

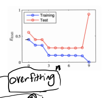


$e^T e = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$

$\hat{e} = y - \hat{y}$

$$\boxed{E(x) = \frac{1}{2} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2}$$

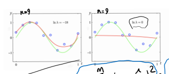




overfitting

$$E(x) = \frac{1}{2} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \frac{\lambda}{2} \|x\|^2$$

$\|x\|^2 = x^T x = a_0^2 + a_1^2 + \dots + a_n^2$



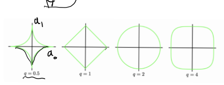
$$E(x) = \frac{1}{2} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \frac{\lambda}{2} \sum_{j=1}^{n} |x_j|^q$$

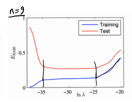
$q=1 \rightarrow$ lasso

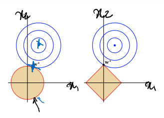
$q=0.5$

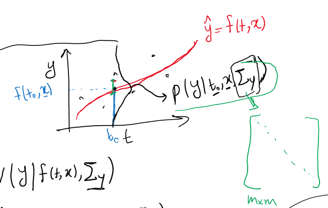
$\boxed{q=2} \rightarrow$ weight Decay

$\hat{x} = (A^T A + \lambda I) A^T y$









likelihood
prior
↓ posterior

Likelihood

$p(y|t, x, \Sigma_y^{-1}) = \mathcal{N}(y|f(t,x), \Sigma_y)$

$p(y|t, x, \Sigma_y^{-1}) = \prod_{i=1}^{m} \mathcal{N}(y_i|f(t_i,x), \Sigma_y)$

$N = f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$

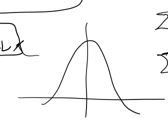
② $\ln p(y|t, x, \Sigma_y^{-1}) = -\dfrac{\Sigma_y^{-1}}{2} \sum_{i=1}^{m} \{f(t_i, x) - y_i\}^2 + \dfrac{m}{2} \ln \Sigma_y^{-1} - \dfrac{m}{2} \ln(2\pi)$

$-\dfrac{1}{2} \sum_{i=1}^{m} \{f(t_i, x) - y_i\}^2$

$E(x) = \dfrac{1}{2} \sum_{i=1}^{m} \{f(t_i, x) - y_i\}^2$

$x_{ML} = (A^T A)^{-1} A^T y$

$\Sigma_{y_{ML}}$



Σ_y

$\Sigma_{y_{ML}}$

$\dfrac{\partial \ln p(y|t, x, \Sigma_y^{-1})}{\partial \Sigma_{y_{ML}}^{-1}}$ $\frac{1}{2}\sum_{i=1}^{m} \{f(t_i, x_{ML}) - y_i\}^2 = \dfrac{m}{2} \dfrac{1}{\Sigma_{y_{ML}}^{-1}}$

$\Sigma_{y_{ML}} = \dfrac{1}{m} \sum_{i=1}^{m} \{f(t_i, x_{ML}) - y_i\}^2$

$p(y|t, x_{ML}, \Sigma_{y_{ML}}^{-1}) = \mathcal{N}(y|f(t, x_{ML}), \Sigma_{y_{ML}})$

prior

$p(x)$ ① $p(x|\Sigma_x^{-1}) = \mathcal{N}(x|0, \Sigma_x I) = \left(\dfrac{\Sigma_x^{-1}}{2\pi}\right)^{\frac{n+1}{2}} \exp\left\{-\dfrac{\Sigma_x^{-1}}{2} x^T x\right\}$

posterior

$p(x|t, y, \Sigma_x^{-1}, \Sigma_y^{-1}) \propto p(y|t, x, \Sigma_y^{-1})\, p(x|\Sigma_x^{-1})$

Solution is MAP → maximum A posteriori

maximize $-\dfrac{\Sigma_y^{-1}}{2} \sum_{i=1}^{m} \{f(t_i, x) - y_i\}^2 - \dfrac{\Sigma_x^{-1}}{2} x^T x$

minimize $\dfrac{\Sigma_y^{-1}}{2} \sum_{i=1}^{m} \{f(t_i, x) - y_i\}^2 + \dfrac{\Sigma_x^{-1}}{2} x^T x$

$\|x\|^2$

$$\boxed{\lambda = \frac{\Sigma_x^{-1}}{\Sigma_y^{-1}} = \frac{\Sigma_y}{\Sigma_x}}$$

$$\boxed{E(x) = \frac{1}{2} \sum_{i=1}^{m} \{f(t_i, x) - y_i\}^2 + \frac{\lambda}{2} \|x\|^2}$$

$a_0 \quad a_1 \quad a_2 \quad \dots$

$\Sigma_x = \begin{bmatrix} 0.5 & & & \\ & 0.3 & & \\ & & 0.2 & \\ & & & \ddots \end{bmatrix}$

Conjugate prior

$p(x|t, y) = \mathcal{N}(x|\mu, \Lambda^{-1})$

$\Lambda^{-1} = \Sigma_y^{-1} A^T A + \Sigma_x^{-1}$

$\mu = \Sigma_y^{-1} \Lambda A^T y$

$\Sigma_y^{-1} (\Sigma_y^{-1} A^T A + \Sigma_x^{-1})^{-1} A^T y$
$(A^T A + \Sigma_x^{-1} \Sigma_y^{-1})^{-1} A^T y$
$(A^T \Sigma_y^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_y^{-1} y$

MAP ← $\mu = (A^T \Sigma_y^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_y^{-1} y$

$\hat{x} = (A^T \Sigma_y^{-1} A)^{-1} A^T \Sigma_y^{-1} y$

prior $\bar{x} = 0$

$p(x|\Sigma_x^{-1}) = \mathcal{N}(x|\mu_0, \Sigma_x^{-1} I)$

MAP $\hat{x} = \mu = (A^T \Sigma_y^{-1} A + \Sigma_x^{-1})^{-1} (A^T \Sigma_y^{-1} y + \Sigma_x^{-1} \mu_0)$

$D(x|y) = (A^T \Sigma_y^{-1} A + \Sigma_x^{-1})^{-1}$

$\hat{x} = (A^T \Sigma_y^{-1} A)^{-1} A^T \Sigma_y^{-1} y$

$D(x|y) = (A^T \Sigma_y^{-1} A)^{-1}$

$\hat{\sigma}_0^2 = \dfrac{e^T \Sigma_y^{-1} e}{m-n}$

$\hat{Q}_{\hat{x}} =$