~~Bagging Example~~

Saturday, 4 March 2023 9:57 AM

$f_i(x) = f^*(x) + \varepsilon_i(x)$

$$\begin{cases} \cdot\ \mathbb{E}\,\varepsilon_i(x) = 0, \quad \mathbb{E}\,\varepsilon_i(x)^2 = \sigma(x)^2 \\ \cdot\ \mathbb{E}\,\varepsilon_i(x)\varepsilon_j(x) = 0 \quad \text{if } i \neq j. \end{cases}$$

Aggregate model: $\bar{f}(x) = \sum_{i=1}^{m} f_i(x)$.

Compare ① $E(x) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\left(f_i(x) - f^*(x)\right)^2$

② $\bar{E}(x) = \mathbb{E}\left(\bar{f}(x) - f^*(x)\right)^2$

① $E(x) = \frac{1}{m}\sum_{i=1}^{m}(\varepsilon_i(x))^2$

$= \frac{1}{m}\sum_{i=1}^{m}\sigma(x)^2 = \sigma(x)^2$

② $\bar{E}(x) = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}\left(f_i(x) - f^*(x)\right)\right]^2$

$= \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}\varepsilon_i(x)\right]^2$

$= \frac{1}{m^2}\sum_{i,j=1}^{m}\mathbb{E}\left(\varepsilon_i(x)\varepsilon_j(x)\right)$

(uncorrelated) $= \frac{1}{m^2}\sum_{i=1}^{m}\sigma(x)^2 = \frac{1}{m}\sigma(x)^2 = \frac{1}{m}E(x)$

Dropout for Linear Regression

Linear model with dropout $f(x; w, \mu) = \frac{1}{p} w^T(\overset{\text{new}}{\mu \circ x})$ $\qquad \mu_j = \begin{cases} 1 & \text{with prob } p \\ 0 & \text{with prob } 1-p. \end{cases}$

Empirical risk (averaged over μ):

$$\hat{R}(w) = \mathbb{E}_\mu \frac{1}{2}\|\underbrace{X\left[\frac{1}{p}D_\mu\right]w}_{f(x;w,\mu)} - y\|^2, \quad D_\mu = \begin{pmatrix} \mu_1 & & 0 \\ & \mu_2 & \\ 0 & & \ddots \\ & & & \mu_d \end{pmatrix}$$

$$= \mathbb{E}_\mu \frac{1}{2} \| \underbrace{Xw - y}_{\text{usual loss}} + \underbrace{X\left[\frac{1}{p} D_\mu - I\right] w}_{\text{new term from dropout}} \|^2$$

$$= \mathbb{E}_\mu \frac{1}{2} \|Xw - y\|^2 + \mathbb{E}_\mu \frac{1}{2} \| X \left[\frac{D_\mu}{p} - I\right] w \|^2$$

$$+ \mathbb{E}_\mu (Xw - y)^T X \underbrace{\left[\frac{1}{p} D_\mu - I\right]}_{\mathbb{E}_\mu = 0} w \quad 0$$

→ cross term repeated twice (scalar)

$$= \frac{1}{2} \|Xw - y\|^2 + \mathbb{E}_\mu \frac{1}{2} \| X \underbrace{\left[\frac{D_\mu}{p} - I\right]}_{R = R(p)} w \|^2$$

$$\mathbb{E}_\mu \|XRw\|^2 = w^T \mathbb{E}_\mu (R^T \underbrace{X^T X}_{d \times d} R) w$$

$$R_{ij} = \delta_{ij} \left(\frac{\mu_i}{p} - 1\right)$$

Let A be a $d \times d$ matrix

$$\left[\mathbb{E}_\mu (R A R)\right]_{ij} = \mathbb{E}_\mu \sum_{k,l} \left(\frac{\mu_i}{p} - 1\right) \delta_{ik} A_{kl} \left(\frac{\mu_l}{p} - 1\right) \delta_{lj}$$

$$= A_{ij} \boxed{\mathbb{E}_\mu \left(\frac{\mu_i}{p} - 1\right)\left(\frac{\mu_j}{p} - 1\right)} \xrightarrow{\text{exercise}} \delta_{ij} \left(\frac{1}{p} - 1\right)$$

$$= A_{ij} \cdot \delta_{ij} \left(\frac{1}{p} - 1\right)$$

Set $A = X^T X$

$$\mathbb{E}_\mu \frac{1}{2} \|XRw\|^2 = \frac{1}{2} \left(\frac{1}{p} - 1\right) w^T \begin{pmatrix} (X^T X)_{11} & & & \\ & (X^T X)_{22} & & \\ & & \ddots & \\ & & & (X^T X)_{dd} \end{pmatrix} w$$

$$= \frac{1}{2} \left(\frac{1}{p} - 1\right)^2 \sum_{i=1}^{d} (X^T X)_{ii} w_i^2$$