Machine Learning 1

Exercise sheet 09

Jinxi Wang 587032 Lusine Nazaretyan 513624

Thomas Herold 535025 Ya Qian 518902

Karen Nazaretyan

Exercise 1.

(a) $\mathrm{Bias}(\hat{\mu}) = E[\hat{\mu} - \mu] = E[\frac{1}{N}\Sigma_{i=1}^N X_i - \mu] = \frac{1}{N} E[\Sigma_{i=1}^N X_i] - \mu$

$= \frac{1}{N} \Sigma_{i=1}^N E[X_i] - \mu$

$= \frac{1}{N} \cdot N \cdot \mu - \mu$

$= 0$

$\mathrm{Var}(\hat{\mu}) = E[(\hat{\mu} - E(\hat{\mu}))^2] = E[(\frac{1}{N}\Sigma_{i=1}^N X_i - \mu)^2]$

$= \frac{1}{N^2} E[(\Sigma_{i=1}^N X_i - N\mu)^2] = \frac{1}{N^2} E[\Sigma_{i=1}^N (X_i - \mu)^2 + \sum_{i \neq j} (X_i - \mu)(X_j - \mu)]$

$= \frac{1}{N^2} \sum_{i=1}^N E(X_i - \mu)^2 + \frac{1}{N^2} \sum_{i \neq j} E(X_i - \mu)(X_j - \mu)$

$\downarrow$ because of independence

$= \frac{1}{N^2} \cdot N \cdot \sigma^2 + \frac{1}{N^2} \sum_{i \neq j} E(X_i - \mu) \cdot E(X_j - \mu)$

$= \frac{1}{N} \sigma^2 + 0 = \frac{\sigma^2}{N}$

$\mathrm{Error}(\hat{\mu}) = \mathrm{Bias}(\hat{\mu})^2 + \mathrm{Var}(\hat{\mu}) = 0^2 + \frac{\sigma^2}{N} = \frac{\sigma^2}{N}$

(b). $\mathrm{Bias}(\hat{\mu}) = E(\hat{\mu} - \mu) = E(0 - \mu) = -\mu.$

$\mathrm{Var}(\hat{\mu}) = E[(\hat{\mu} - E(\hat{\mu}))^2] = E[(0 - E(0))^2] = 0.$

$\mathrm{Error}(\hat{\mu}) = \mathrm{Bias}(\hat{\mu})^2 + \mathrm{Var}(\hat{\mu}) = \mu^2$

Exercise 2.

(a). $\text{Error}(\hat{f}(x)) = E[(\hat{f}(x) - f(x))^2]$, $\text{Bias}(\hat{f}(x)) = E[\hat{f}(x) - f(x)]$,

$\text{Var}(\hat{f}(x)) = E[(\hat{f}(x) - E[\hat{f}(x)])^2]$

$$
\begin{aligned}
\text{Error}(\hat{f}(x)) &= E[(\hat{f}(x) - f(x))^2] \\
&= E[(\hat{f}(x) - E(\hat{f}(x)) + E(\hat{f}(x)) - f(x))^2] \\
&= E\{[\hat{f}(x) - E(\hat{f}(x))]^2 + 2[\hat{f}(x) - E(\hat{f}(x))][E(\hat{f}(x)) - f(x)] + [E(\hat{f}(x)) - f(x)]^2\} \\
&= E[(\hat{f}(x) - E[\hat{f}(x)])^2] + 2E[\hat{f}(x) - E(\hat{f}(x))][E(\hat{f}(x)) - f(x)] + E[E(\hat{f}(x)) - f(x)]^2 \\
&= \text{Var}(\hat{f}(x)) + 2\times 0 + [E(\hat{f}(x)) - f(x)]^2 \quad (\because E(E(\hat{f}(x))) = E(\hat{f}(x)), E(f(x)) = f(x)) \\
&= \text{Var}(\hat{f}(x)) + \text{Bias}(\hat{f}(x))^2
\end{aligned}
$$

Exercise 3.

(a). $D_{KL}(R\|\hat{p}) = \sum_{i=1}^{c} R(i) \log \frac{R(i)}{\hat{p}(i)}$

$$\min_R E[D_{KL}(R\|\hat{p})] = \min_R [\sum_{i=1}^{c} R(i) \log \frac{R(i)}{\hat{p}(i)}]$$

s.t. $\sum_{i=1}^{c} R(i) = 1$.

$$\mathcal{L} = \sum_{i=1}^{c} R(i) \log R(i) - \sum_{i=1}^{c} R(i) \log(\hat{p}(i)) + \lambda[\sum_{i=1}^{c} R(i) - 1]$$

$$\frac{\partial \mathcal{L}}{\partial R(i)} = \log R(i) + 1 - \log(\hat{p}(i)) + \lambda = 0 \quad (*), \quad i = 1, 2, \cdots, c.$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{c} R(i) - 1 = 0 \quad (**)$$

From $(*)$, we have $R(i) = \exp[\log(\hat{p}(i)) - \lambda - 1) = \frac{\exp[\log \hat{p}(i)]}{\exp(\lambda+1)}$ $(***)$

Plug it into $(**)$, we can obtain

$$\sum_{i=1}^{c} \exp(\log(\hat{p}(i)))/\exp(\lambda+1) = 1$$

$$\Rightarrow \exp(\lambda+1) = \sum_{i=1}^{c} \exp \log(\hat{p}(i))$$

Plug it back into $(***)$, we have the solution to the optimization problem

$$R(i) = \frac{\exp E[\log \hat{p}(i)]}{\sum_j \exp E \log \hat{p}(j)}$$

(b). $\text{Error}(\hat{p}) = E\left(\sum_{i=1}^{C} P(i)\log\frac{P(i)}{\hat{p}(i)}\right) = \sum_{i=1}^{C} P(i)\log P(i) - \sum_{i=1}^{C} P(i) E[\log \hat{P}(i)]$ (1)

$\text{Bias}(\hat{p}) = D_{KL}(P\|R) = \sum_{i=1}^{C} P(i)\log\frac{P(i)}{R(i)} = \sum_{i=1}^{C} P(i)\log P(i) - \sum_{i=1}^{C} P(i)\log R(i)$ (2)

$\text{Var}(\hat{p}) = E[D_{KL}(R\|\hat{P})] = E\left[\sum_{i=1}^{C} R(i)\log\frac{R(i)}{\hat{p}(i)}\right] = \sum_{i=1}^{C} R(i)\log R(i) - \sum_{i=1}^{C} R(i) E[\log(\hat{p}(i))]$ (3)

From (1), (2), (3), we know to prove

$\text{Error}(\hat{p}) = \text{Bias}(\hat{p}) + \text{Var}(\hat{p})$ is equivalent to prove

$$\sum_{i=1}^{C} P_i \log P(i) - \sum_{i=1}^{C} P(i) E[\log \hat{P}(i)] = \sum_{i=1}^{C} P(i)\log P(i) - \sum_{i=1}^{C} P(i)\log R(i) + \sum_{i=1}^{C} R(i)\log R(i) - \sum_{i=1}^{C} R(i) E[\log(\hat{P}(i))]$$

$\Leftrightarrow$

$$\sum_{i=1}^{C} P(i) E[\log \hat{P}(i)] = \sum_{i=1}^{C} P(i)\log R(i) + \sum_{i=1}^{C} R(i) E[\log \hat{P}(i)] - \sum_{i=1}^{C} R(i)\log R(i) \quad \hookrightarrow (4)$$

After insert $R(i)$ from (a),

The RHS of (4) $= \sum_{i=1}^{C}\left[P(i) - \frac{\exp E[\log \hat{P}(i)]}{\sum_{j=1}^{C}\exp E(\log \hat{P}(j))}\right]\cdot\left[E[\log \hat{P}(i)] - \sum_{j=1}^{C} E(\log \hat{P}(j))\right]$

$$+ \sum_{i=1}^{C} \frac{\exp E[\log \hat{P}(i)]}{\sum_{j=1}^{C}\exp E[\log \hat{P}(j)]}\cdot E[\log \hat{P}(i)]$$

$$= \sum_{i=1}^{C} P(i) E[\log \hat{P}(i)] - \sum_{i=1}^{C}\left[P(i)\sum_{j=1}^{C} E(\log \hat{P}(j))\right]$$

$$\cancel{- \sum_{i=1}^{C} \frac{\exp E[\log \hat{P}(i)]}{\sum_{j=1}^{C}\exp E(\log \hat{P}(j))}\cdot E[\log \hat{P}(i)]} + \sum_{i=1}^{C}\left(\frac{\exp E[\log \hat{P}(i)]}{\sum_{j=1}^{C}\exp E(\log \hat{P}(j))}\cdot \sum_{j=1}^{C} E(\log(\hat{P}(j)))\right)$$

$$\cancel{+ \sum_{i=1}^{C} \frac{\exp E[\log \hat{P}(i)]}{\sum_{j=1}^{C}\exp E[\log \hat{P}(j)]}\cdot E[\log \hat{P}(i)]}$$

$$= \sum_{i=1}^{C} P(i) E[\log \hat{P}(i)] - \cancel{\sum_{j=1}^{C} E[\log \hat{P}(j)]\cdot \underbrace{\sum_{i=1}^{C} P(i)}_{=1}}$$

$$\cancel{+ \sum_{j=1}^{C} E[\log \hat{P}(j)] \underbrace{\sum_{i=1}^{C} \frac{\exp E[\log \hat{P}(i)]}{\sum_{j=1}^{C}\exp E[\log \hat{P}(j)]}}_{=1}}$$

$$= \sum_{i=1}^{C} P(i) E[\log \hat{P}(i)]$$

$=$ LHS of (4).

which proves $\text{Error}(\hat{p}) = \text{Bias}(\hat{p}) + \text{Var}(\hat{p})$