Ruhani Bhatia
2023450

Q.1.

SML Assignment – III

(1)

1) $R(f)$ ↳ Expected Risk $= E[(Y - f(X))^2]$

$$= \int (y - f(x))^2 \underbrace{p(x,y)}\, dx\, dy$$

Bayes Theorem

$$= \int (y - f(x))^2\, p(y|x)\, p(x)\, dx\, dy$$

$$= \int \underbrace{\left[\int (y - f(x))^2 \cdot p(y|x)\, dy\right]}\, p(x)\, dx$$

$= J(f(x))$ conditional expectation of $(Y - f(X))^2$ given $X = x$

minimise $R(f) \Rightarrow$ minimise $\underbrace{\int (y - f(x))^2\, p(y|x)\, dy}_{= J(f(x))}$ with respect to $f(x)$

$$\Rightarrow \frac{\partial J}{\partial f(x)} = \int 2(y - f^*(x))(-1)\, p(y|x)\, dy = 0$$

$$\Rightarrow \int y\, p(y|x)\, dy = f^*(x) \underbrace{\int p(y|x)\, dy}_{=1}$$

$$\Rightarrow \int y\, p(y|x)\, dy = f^*(x) = E[Y | X = x]$$

$f^*(x)$ that minimises the true risk is the conditional expectation

$\Rightarrow f^*(x) = E(Y | X = x]$

2) $f^*(x) = E[Y | X = x]$ represents the conditional mean of output Y given the input $X = x$. In other words, given any input x, the best prediction (in sense of minimising expected square error) is avg value of Y conditioned on $X = x$.

It is the optimal pt estimate ~~___~~ of Y given X under squared loss.

→ $\hat{f}(x)$ captures central tendency of Y given X=x
→ It does not capture other aspects of $p(y|x)$ such as variance & skewness
→ If $p(y|x)$ is symmetric (gaussian or Poisson) $f^*(x)$ is also mode & median of $p(y|x)$
→ if $p(y|x)$ is asymmetric, $f^*(x)$ may not be most probable value (mode), but it still minimises the squared loss.

Q2. 1) $E(\hat{f}(x)) = \left(\frac{1.9+2.1+2}{3}\right)x + \left(\frac{3.5+3.2+3.4}{3}\right)$

$f_{avg}(x) = (2.0)x + \left(\frac{10.1}{3}\right)$

2) $Bias = E(\hat{f}(x)) - f(x) = (-f(2) + E(\hat{f}(2)))$ $\left(\frac{10.1}{3} \approx 3.37\right)$

$= +(2\times(2) + \frac{10.1}{3}) - 2(2)+3$

$= +\frac{1.1}{3} \approx 0.37$

$Bias^2 = (0.37)^2 \approx 0.13444\ldots = \left(\frac{1.1}{3}\right)^2$

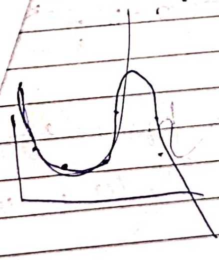
3) $Var_{at\ x=2} = E\left(E(\hat{f}(x)) - \hat{f}(x)\right)^2 \Big|_{at\ x=2}$

$= \frac{1}{3}\sum_{i=1}^{3}\left(f_{avg}(2) - \hat{f}_i(2)\right)^2$

$= \frac{1}{3}[7.37 - 3.8 - 3.5]^2 + \frac{1}{3}[7.37 - 4.2 - 3.2]^2 + \frac{1}{3}[7.37 - 4 - 3.4]^2$

$= \frac{1}{3}\left((0.07)^2 + (-0.03)^2 + (-0.03)^2\right)$

$= \frac{0.0067}{3} = 0.0022333\ldots$



②

Ruhani Bhatia
2023450

4) $E_D(\hat{f}(x) - f(x))^2$

True $MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{f}_i(2) - f(2))^2$

Vs $MSE(x) = bias^2 + Var + \eta^2 = E(\hat{f}(x) - y)^2$
$\approx 0.236677\ldots$ ① (taking bias $= \frac{1.1}{3}$)

if bias = 0.37 ⇒ $MSE_{at\ x=2} = 0.239133\ldots$

$f(2) = 2\times2 + 3 = 7$

$= \frac{1}{3}\left[(\hat{f}_1(2) - 7)^2 + (\hat{f}_2(2) - 7)^2 + (\hat{f}_3(2) - 7)^2\right]$

$= \frac{(0.3)^2 + (0.4)^2 + (0.4)^2}{3}$

$= \frac{0.41}{3} = 0.1367$ ②

① − ② ≈ 0.1 = η^2 → variance of noise (inherent noise in labels)

∴ expected squared error calculated from data pt
$= 0.1367 + 0.1 \approx 0.2367$

which is very close to MSE calculated from bias − variance decomposition ≈ 0.236677

$|0.2367 - 0.236677| \approx 0$ (they are very close in magnitude).