ECE 461P: Homework 2 Scratch work

Sidharth Babu and Laith Altarabishi

February 16, 2023

1 Problem 1

1.1 Bias-Variance Decomposition

Target: $E[(y(x|D) - h(x))^2]$

$$\begin{aligned}(y(x|D) - h(x))^2 &= (y(x|D) - E_d[y(x|D)] + E_d[y(x|D)] - h(x))^2 \\ &= (y(x|D) - E_d[y(x|D)])^2 + (E_d[y(x|D)] - h(x))^2 + 2(y(x|D) - E_d[y(x|D)])(E_d[y(x|D)] - h(x))\end{aligned}$$

Let $y(x|D) = y$ and $h(x) = h$
Taking Expectation of both sides:

$$\begin{aligned}E[(y-h)^2] &= E[(y - E[y])^2] + E[E[y]^2 - 2E[y] + h^2] - 2E[(y - E[y])(E[y] - h)] \\ &= Var(y) + E[y]^2 - 2E[y]E[h] + E[h]^2 - 2E[(y - E[y])(E[y] - h)] \\ &= Var(y) + E[y]^2 - 2E[y]E[h] + E[h]^2 - 2E[y]^2 + 2E[y][h] + 2E[y]^2 - 2E[h]E[y] \\ &= Var(y) + E[y]^2 + E[h]^2 - 2E[y]E[h] \\ &= Var(y) + (E[y] - E[h])^2\end{aligned}$$

For a given X:

$$E[(y(x|D) - h(x))^2] = Var(y(x|D)) + (E[y(x|D)] - h(x))^2$$

2 Problem 2

2.1 Log-Likelihood

Given:

$$\begin{aligned}&y = w_0 x + \epsilon \\ &\epsilon \sim N(0, \sigma^2) if x_i > 0 \\ &\epsilon \sim N(0, 4\sigma^2) if x_i <= 0\end{aligned}$$

Therefore:

$$\begin{aligned}&y \sim N(w_0 x, \sigma^2 | x_i > 0) \\ &y \sim N(w_0 x, 4\sigma^2 | x_i <= 0)\end{aligned}$$

NLL calculation for a constant noise model:

$$\begin{aligned}
\mathcal{L}(Y_n) &= \prod N(y_i|wx_0, \sigma^2) \\
l(Y_n) &= \sum \log N(y_i|wx_0, \sigma^2) \\
&= \sum \log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - wx_0)^2}{2\sigma^2}} \\
&= \sum \log(1) - \log(\sqrt{2\pi\sigma^2}) + \log(e^{-\frac{(y_i - wx_0)^2}{2\sigma^2}}) \\
&= \sum -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y_i - wx_0)^2}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum (y_i - wx_0)^2
\end{aligned}$$

NLL calculation for a non-constant noise model:

$$\begin{aligned}
\mathcal{L}(Y_n) &= \prod_{i:x_i>0} N(y_i|wx_0, \sigma^2) \prod_{i:x_i<=0} N(y_i|wx_0, 4\sigma^2) \\
l(Y_n) &= \sum_{i:x_i>0} \log N(y_i|wx_0, \sigma^2) + \sum_{i:x_i<=0} \log N(y_i|wx_0, 4\sigma^2) \\
&= \sum_{i:x_i>0} \log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - wx_0)^2}{2\sigma^2}} + \sum_{i:x_i<=0} \log \frac{1}{\sqrt{2\pi 4\sigma^2}} e^{-\frac{(y_i - wx_0)^2}{2(4\sigma^2)}} \\
&= \sum_{i:x_i>0} \log(1) - \log(\sqrt{2\pi\sigma^2}) + \log(e^{-\frac{(y_i - wx_0)^2}{2\sigma^2}}) + \sum_{i:x_i<=0} \log(1) - \log(\sqrt{8\pi\sigma^2}) + \log(e^{-\frac{(y_i - wx_0)^2}{8\sigma^2}}) \\
&= \sum_{i:x_i>0} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y_i - wx_0)^2}{2\sigma^2} + \sum_{i:x_i<=0} -\frac{1}{2}\log(8\pi\sigma^2) - \frac{(y_i - wx_0)^2}{8\sigma^2} \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i:x_i>0}(y_i - wx_0)^2 - \frac{n}{2}\log(8\pi\sigma^2) - \frac{1}{8\sigma^2}\sum_{i:x_i<=0}(y_i - wx_0)^2 \\
&= -\frac{n}{2}(\log(2\pi\sigma^2) + \log(8\pi\sigma^2)) - \frac{1}{2\sigma^2}\sum_{i:x_i>0}(y_i - wx_0)^2 - \frac{1}{8\sigma^2}\sum_{i:x_i<=0}(y_i - wx_0)^2 \\
&= -\frac{n}{2}(\log(2\pi\sigma^2 * 8\pi\sigma^2)) - \frac{1}{2\sigma^2}\sum_{i:x_i>0}(y_i - wx_0)^2 - \frac{1}{8\sigma^2}\sum_{i:x_i<=0}(y_i - wx_0)^2 \\
&= -\frac{n}{2}\log(16\pi\sigma^4) - \frac{1}{2\sigma^2}\sum_{i:x_i>0}(y_i - wx_0)^2 - \frac{1}{8\sigma^2}\sum_{i:x_i<=0}(y_i - wx_0)^2
\end{aligned}$$

3 Problem 4

The one-step update equation for w is:

$$
\begin{aligned}
w_1 &= w_0 - \eta(\nabla E) \\
&= w_0 - \eta(\frac{\partial E}{\partial w}(\frac{1}{N}\sum_{i}^{N}(y_i - w^T x_i)^2))
\end{aligned}
$$

Since it is SGD, N is 1:

$$
\begin{aligned}
&= w_0 - \eta(\frac{\partial E}{\partial w}((y_i - w^T x_i)^2)) \\
&= w_0 - \eta(-2x_i(y_i - w^T x_i)) \\
&= w_0 + \eta(2x_i(y_i - w^T x_i))
\end{aligned}
$$