Problem 1

Lemma: For any model, the bias-var decomposition is given by

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}\left[\hat{f}_K^{(\mathcal{D})}(x_0)-y_0)^2\right] &= \mathbb{E}_{\mathcal{D}}\left[(\hat{f}_K^{(\mathcal{D})}(x_0)-\bar{f}(x_0)+\bar{f}(x_0)-f(x_0)-\epsilon)^2\right] \\ &= \mathbb{E}_{\mathcal{D}}\left[(\hat{f}_K^{(\mathcal{D})}(x_0)-\bar{f}(x_0))^2\right]+\mathbb{E}\left[(\bar{f}(x_0)-f(x_0))^2\right]+\sigma^2\end{aligned} \tag{1}$$

The equation above hold because

$$\mathbb{E}(\epsilon)=0 \qquad \epsilon^2=\sigma \tag{2}$$

$$\mathbb{E}_{\mathcal{D}}\left[\hat{f}_K^{(\mathcal{D})}(x_0)-\bar{f}(x_0)\right]=0 \tag{3}$$

The cross-product are hence all zero.
In the special case of kNN estimation

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}\left[(\hat{f}_K^{(\mathcal{D})}(x_0)-y_0)^2\right] &= \mathbb{E}_{\mathcal{D}}\left[(\hat{f}_K^{(\mathcal{D})}(x_0)-\bar{f}(x_0))^2\right]+bias^2+\sigma^2 \\ &= \mathbb{E}_{\mathcal{D}}\left\{\frac{1}{K}\sum_{i=1}^{K}y_i-\mathbb{E}\left[\frac{1}{K}\sum_{i=1}^{K}y_i\right]\right\}+bias^2+\sigma^2 \\ &= \frac{1}{K}Var(y)+bias^2+\sigma^2 \\ &= bias^2+\frac{\sigma^2}{K}+\sigma^2\end{aligned} \tag{4}$$

The last equal sign holds because

$$Var(y)=Var(\epsilon)=\sigma^2 \tag{5}$$

Problem 3

WMSE can be written into matrix form

$$WMSE=(X\beta-Y)^TW(X\beta-Y) \tag{6}$$

where W is the diagonal matrix with entries of weights

$$\hat{\beta}=\arg\min_{\beta}\left\{(X\beta-Y)^TW(X\beta-Y)\right\} \tag{7}$$

F.O.C.

$$\nabla_{\hat{\beta}}\{\beta^TX^TWX\beta-\beta^TX^TWY-Y^TWX\beta+Y^TWY\}=0 \tag{8}$$

$$\begin{aligned}X^TWX\hat{\beta}&=X^TWY \\ \hat{\beta}&=(X^TWX)^{-1}X^TWY\end{aligned} \tag{9}$$

Problem 4

The log-likelihood function is

$$\begin{aligned} l(b) &= \sum_{i=1}^{n} y_i ln(h(x_i)) + \sum_{i=1}^{n}(1-y_i)ln(1-h(x_i)) \\ &= \sum_{i=1}^{n} y_i ln(\frac{e^{b^T x_i}}{1+e^{b^T x_i}}) + \sum_{i=1}^{n}(1-y_i)ln(\frac{1}{1+e^{b^T x_i}}) \end{aligned} \tag{10}$$

The gradient of $l(b)$ is

$$\begin{aligned} \nabla l(b) &= \sum_{i=1}^{n} \left\{ y_i \frac{1+e^{b^T x_i}}{e^{b^T x_i}} \frac{x_i e^{b^T x_i}}{(1+e^{b^T x_i})^2} + (1-y_i)(1+e^{b^T x_i}) \frac{-x_i e^{b^T x_i}}{(1+e^{b^T x_i})^2} \right\} \\ &= \sum_{i=1}^{n} \{ y_i x_i (1-h(x_i)) - (1-y_i) x_i h(x_i) \} \\ &= \sum_{i=1}^{n} \left\{ x_i y_i - x_i \phi(b^T x_i) - x_i y_i \phi(b^T x_i) + x_i y_i \phi(b^T x_i) \right\} \\ &= \sum_{i=1}^{n} \left\{ (y_i - \phi(b^T x_i)) x_i \right\} \end{aligned} \tag{11}$$

Problem 5

The log-likelihood function is

$$\begin{aligned} \sum_{i=1}^{n} ln(\phi(y_i b^T x_i)) &= \sum_{i=1}^{n} \left\{ ln(\frac{e^{y_i b^T x_i}}{1+e^{y_i b^T x_i}}) \right\} \\ &= \sum_{i=1}^{n} \left\{ -ln(\frac{1+e^{y_i b^T x_i}}{e^{y_i b^T x_i}}) \right\} \\ &= \sum_{i=1}^{n} \left\{ -ln(1+e^{-y_i b^T x_i}) \right\} \\ &= n E_{in}(b) \end{aligned} \tag{12}$$

Take the gradient w.r.t. b

$$\nabla E_{in}(b) = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i y_i e^{y_i b^T x_i}}{1+e^{-y_i b^T x_i}} \tag{13}$$

$$\nabla E_{in}(b) = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i y_i}{1+e^{y_i b^T x_i}} \tag{14}$$