COMS 4771 Fall 2016 Homework 5 solutions (partial)

Problem 2(a)

As usual, we look at the conditional expectation

$$\mathbb{E}[\ell(Yf(X)) \mid X = x] \;=\; \mathbb{E}[e^{-Yf(x)} \mid X = x] \;=\; \eta(x)e^{-f(x)} + (1-\eta(x))e^{f(x)}$$

for an arbitrary $x \in \mathcal{X}$. This is a convex function of $f(x)$, and its derivative with respect to $f(x)$ is

$$\eta(x) \cdot (-1)e^{-f(x)} + (1-\eta(x)) \cdot e^{f(x)} .$$

We find that the derivative is equal to zero when

$$f(x) \;=\; \frac{1}{2} \ln \frac{\eta(x)}{1-\eta(x)} .$$

Hence, the conditional expectation is minimized at this value of $f(x)$. We conclude that the function f that minimizes $\mathbb{E}[\ell(Yf(X))]$ is half the log-odds ratio function.

Problem 2(c)

It turns out the conditional expectation function was actually an affine function. So using square loss instead of logistic loss would give a better MAE.

Problem 3(a)

1. True ($y_i \mid x_i \sim P_{(w_\star, \sigma_\star^2)}$)
2. False ($w_\star$ is not random)
3. True (from lecture)
4. True (since $\hat{w}_{\mathrm{ols}}$ is unbiased estimator of $w_\star$)
5. True (if Π is orthogonal projector to range of A, then $I - \Pi$ is orthogonal projector to left null space of A; $(I-\Pi)y = y - \Pi y = y - A\hat{w}_{\mathrm{ols}}$)
6. True ($\sum_{i=1}^n r_i = 0$ means r is orthogonal to the all-ones vector, which is the first column of A; we know r is orthogonal to every column of A)