Statistical Learning I - Homework based on August 24 and 29 lectures.

1. Recall in class, we showed that with squared error loss, the best predictor is the *regression function* $f(x) = E[Y \mid X = x]$. Of course to find this function exactly, we would have to know how X and Y are jointly distributed. In this exercise, we will get some practice calculating the regression function when the joint distribution of X and Y is known. Suppose for now that X and Y are random variables with joint `pdf` $p(x, y) = e^{-y}$ for $0 < x < y < +\infty$. We wish to predict Y from X. Let's find the regression function. We will do this in steps. Be careful of the region on which the joint density is defined!

- **a.** First find the marginal density of X using $p(x) = \int p(x, y)\, dy$.
- **b.** Now find the conditional density of Y given X using $p(y \mid x) = \frac{p(x,y)}{p(x)}$.
- **c.** Now calculate the regression function $E[Y \mid X = x] = \int y\, p(y \mid x)\, dy$.

2. Let's see how well we calculated the regression function in the previous problem. I generated some data using the joint density above and it is contained in a file called *mydata.csv* on OAKS. Load this data into R and plot the (x, y) pairs. Next construct a simple linear model using the `lm` command in R. Look at the summary and plot the line on the same plot as the data using the `abline` command. Does this line roughly match the *true* line that you calculated in the first exercise?

3. Consider the linear model $Y = X\beta + \epsilon$ where ϵ is random error with mean zero. In simple linear regression, Y is an n-vector, X is an $n \times 2$ matrix with the first column being all ones, $\beta = (\beta_0, \beta_1)$, and ϵ is an n-vector. In class we showed that the values of the β coefficients that minimized the residual sums of squares $RSS(\beta) = (Y - X\beta)^T(Y - X\beta) = Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta$ are given by $\hat{\beta} = (X^TX)^{-1}X^TY$. For this simple linear regression situation, show that

- **a.** $d\,\beta^TX^TY/d\,\beta = X^TY$.
- **b.** $d\,\beta^TX^TX\beta/d\,\beta = 2X^TX\beta$.

4 a. Show that $I - H$ is a projection matrix.

- **b.** We know that $Y^TY = \|Y\|^2$. Now $Y^TY = Y^T(I - H + H)Y$. Use this expression and the properties of projection matrices to show that $\|Y\|^2 = \|\hat{Y}\|^2 + \|Y - \hat{Y}\|^2$.

5a. Load the *Body Fat* data and consider the model that predicts body fat from the other variables. Letting X be the design matrix for this model, calculate $(X^TX)^{-1}$ in R and multiply this matrix by the estimate $\hat{\sigma}^2$. You will need construct the linear model to get this. Now compare the standard errors for the estimates in the summary of the linear model with the square roots of the diagonals of $\hat{\sigma}^2(X^TX)^{-1}$.

- **b.** Consider all of the variables that were not significant at the $\alpha = 0.05$ level in the full body fat model. Can they all be removed? Construct an F-test in R to determine whether or not they can be removed.