# MATH 390.03-02 / 650 Fall 2015 Homework #8

## Professor Adam Kapelner

## Due 4PM Thursday, April 21, 2016

(this document last updated Thursday 14th April, 2016 at 10:08pm)

### Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out." Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read about the normal-normal conjugate and semi-conjugate model and univariate regression. Also read ch15 in McGrayne.

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]" and purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

Problems marked "[MA]" are for the masters students only (those enrolled in the 650 course). For those in 390, doing these questions will count as extra credit.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 10 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com`. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: _____

## Problem 1

This problem is about the normal-normal model using a "semi-conjugate" prior. Reference relevant questions in HW7.

(a) [easy] If $\theta$ and $\sigma^2$ are assumed to be independent, how can $\mathbb{P}(\theta, \sigma^2)$ be factored?

(b) [easy] If $\theta \sim \mathcal{N}(\mu_0, \tau^2)$ and $\sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$, find the kernel of $\mathbb{P}(\theta, \sigma^2)$. (Note that I have renamed $\nu_0$ to $n_0$ to reflect that this hyperparameter is directly interpretable as number of prior samples you've seen in the normal-normal conjugate model but if you use the old notation, I will not penalize you.)

(c) [easy] Using your answer to (b), find the kernel of $\mathbb{P}(\theta, \sigma^2 \mid X)$.
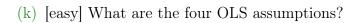
(d) [difficult] Show that the kernel in (c) cannot be factored into the kernel of a normal and the kernel of an inverse gamma. This is lecture 15 bottom of page 1. My algebra may not be 100% correct there.

(e) [difficult] Your answer to (d) looks like a normal and a $k(\sigma^2 \mid X)$. Find its mode.

(f) [easy] Describe how you would sample from $\mathbb{P}\left(\sigma^2 \mid X\right)$ using a normal approximation.

(g) [easy] Describe how you would sample from $\mathbb{P}\left(\sigma^2 \mid X\right)$ using a grid approximation.

(h) [difficult] What's the bad part about not using a conjugate model?

## Problem 2

This problem is about least squares regression.

(a) [easy] What are some names for the $x$ variable?

(b) [easy] What are some names for the $y$ variable?

(c) [easy] How is this data different than all other data we've considered thus far in the semester?

(d) [easy] How are $x$ and $y$ related? Use the general $f$ for the functional relationship here.

(e) [easy] What is $\mathcal{E}$?

(f) [easy] Why should we limit the size of $f$?

(g) [easy] Describe the set $\mathcal{F}_\ell$, the set of all linear models.

(h) [easy] What is the objective function to minimize when fitting a linear model?

(i) [difficult] Consider you have fit a univariate linear regression and have computed the least squares estimates $b_0$ and $b_1$. Define the residual $e_i = y_i - (b_0 + b_1 x_i)$ i.e the distance between the true value of the response and your predicted value using the best fit line. Does $e_i = \epsilon_i$?

(j) [difficult] Prove that the average of the residuals is zero always.

(k) [easy] What are the four OLS assumptions?

(l) [harder] Why is normality of the errors a reasonable thing to assume?

(m) [easy] Assume the four OLS assumptions. What are the MLE's for $\beta_0$ and $\beta_1$? No need to derive them; just copy them from the notes.

(n) [harder] Explain why $b_1 \neq \beta_1 \neq B_1$.

(o) [harder] In regression, the conditional expectation function is called $\mathbb{E}[Y \mid X]$. Find its value when assuming the OLS assumptions.

These are questions about logistic regression.

(a) [harder] What is the distribution of $Y$ under the logistic linear model. You cannot use the $\mathbb{P}(\cdot)$ function in your answer.

(b) [difficult] [MA] Why is there no closed form solution for $b_0$ and $b_1$?

(c) [easy] Consider $b_0 = 3$ and $b_1 = -1$. What is the probability of $Y$ being the positive classs if $x = 3$? If your answer to (a) is correct, this is simple.

(d) [easy] Consider $b_0 = 3$ and $b_1 = -1$. What is the probability of $Y$ being the positive classs if $x = 5.3$? If your answer to (a) is correct, this is simple.

(e) [harder] Consider $b_0 = 3$ and $b_1 = -1$. What is the probability of $Y$ being the positive class if $x = 1000$? Is your calculator lying to you?

## Problem 4

These are questions about the multivariate regression.

(a) [easy] Under the OLS assumptions, what is the distribution of $Y \mid \boldsymbol{X}$?

(b) [harder] What is the PDF of the r.v. from (a)?

(c) [easy] Is each $Y_i \overset{iid}{\sim}$ ?

(d) [harder] Is each $Y_i \mid X_i \overset{iid}{\sim}$ ?

(e) [harder] [MA] Find the distribution of $\boldsymbol{X}^T Y \mid \boldsymbol{X}$.

(f) [difficult] [MA] Under the OLS assumptions, prove that the MLE for $\beta$ (the vector) is the same as the least squares solution $b = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^Ty$.

(g) [easy] Under the OLS assumptions, what is the distribution of $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^TY \mid \boldsymbol{X}$?

(h) [harder] What is the PDF of the r.v. from (g)? Expand all terms.

(i) [difficult] [MA] Is $B_1$ independent of $B_9$? Explain why or why not.

(j) [difficult] [MA] How would you calculate $\mathbb{V}\mathrm{ar}\,[B_1 + B_9]$?

## Problem 5

These are questions about ridge for univariate regression.

(a) [easy] Write the objective function we are minimizing with the $L2$ penalty (end of Lecture 18). Use the notation $m$ for the penalty on $\beta_0^2 + \beta_1^2$.

(b) [easy] Write the objective function from (a) in terms of $y_i$, $x_i$, $\beta_0$, $\beta_1$ and $m$ and standard math operators.

(c) [difficult] Find the ridge estimator for $b_{R,1}$ by minimizing the objective function in (b) as a function of $\beta_1$. You may need to work on the next problem a bit too.

(d) [difficult] Find the ridge estimator for $b_{R,0}$ by minimizing the objective function in (b) as a function of $\beta_0$. You may need to work on the previous problem a bit too.

(e) [difficult] Find the kernel of $\mathbb{P}\left(\beta_1, \beta_0 \mid X, y, \sigma^2\right)$ assuming the OLS assumptions and the priors $\beta_0 \sim \mathcal{N}\left(0, \sigma^2/m\right)$ and $\beta_1 \sim \mathcal{N}\left(0, \sigma^2/m\right)$.

(f) [difficult] Find the mode of the kernel in (e) in both the $\beta_0$ and $\beta_1$ dimensions simultaneously. Hint: you may already have derived this.

(g) [difficult] What is the shrinkage factor $\rho$ on the ridge estimate for $\beta_0$?

(h) [difficult] What is the shrinkage factor $\rho$ on the ridge estimate for $\beta_1$?

(i) [easy] What does the ridge estimate for $\beta_1$ become as $m$ becomes really small?

(j) [easy] What does the ridge estimate for $\beta_1$ become as $m$ becomes really large?