STAT 391
Homework 8
Out Wednesdy May 28, 2020
Due Wednesday June 3, 2020

©Marina Meilă
mmp@stat.washington.edu

Problem 1 – Least Squares

Let $x_1, x_2, \ldots x_n$ be real numbers, and define by $g(z)$ the function

$$g(z) \;=\; \sum_{i=1}^{n}(x_i - z)^2$$

Show that the minimum of g is attained for

$$z^* \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i$$

What is the value $g(z^*)$?

[Hint: Take the derivative of g w.r.t. z and solve the equation $g'(z) = 0$.]

Problem 2 – Logistic regression

Notations follow the textbook.

a. Prove that $\sigma(u) + \sigma(-u) = 1$ for all u, where σ is the logistic CDF/sigmoid function.

b. Prove that if $u \to \pm\infty$, the derivative $\sigma'(u) \to 0$. Hence, correctly classified data points away from the boundary have little influence on the log-likelihood. In other words, the parameters b, γ will change little, when points far away from the the decision boundary move (or appear/disappear).

Problem 3 – Penalized regression

For this problem, assume $x, \beta \in \mathbb{R}$. Assume that $p(y_{1:n}|x_{1:n}, \beta)$ is the usual Normal model for Least Squares linear regression.

a. *Ridge regression* optimizes the cost function

$$J_\lambda(\beta) \;=\; \sum_{i=1}^{n}(y_i - \beta x_i)^2 + \lambda\beta^2, \tag{1}$$

where $\lambda > 0$ is a *regularization parameter. This parameter is like a smoothing parameter in kernel density estimation, in the sense that it is fixed before we see the data and estimate β.*

Find the analytic solution β^{ridge} of (1) as a function of β^{ML} the Least Squares ML estimate of β.

b. Show that that $\beta^{\text{ridge}} < \beta^{ML}$ for all $\lambda > 0$.

[**c.** – **Extra credit**] Show that one can augment the data $(y_i, x_i)_{1:n}$ so that β^{ridge} is the ML estimate for the new data. Write the corresponding statisical model $P(y^{new}|x^{new})$.

[**d.** – **Extra credit**] Now consider Bayesian estimation. Find a prior $p_0(\beta)$ so that

$$J_\lambda(\beta) \;=\; A \ln p(\beta \,|\, y_{1:n}, x_{1:n}) + B, \tag{2}$$

where A, B are constants independent on β. You need not determine the value of B.