Statistical Learning - Milan, Fall2019

Homework problem 4

Ways of interpreting and calculating Ridge and Lasso regression:

1. **ESL 3.7:** Show that if we assume a likelihood $y_i \sim N(x_i^T\beta, \sigma^2)$ for $i = 1, ..., n$ and a prior $\beta \sim N(0, \tau^2 I)$, then the negative log-posterior density of β is $\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$ up to multiplication and addition of constants, with $\lambda = \sigma^2/\tau^2$. Conclude that the Ridge solution is a maximum posterior estimate of β.

2. Show that the same applies to Lasso, except that the prior on β is a double exponential.
   Note: A double exponential random variable with parameter θ has density $f(x) = \theta/2 \cdot \exp(-|x|\theta)$.

3. **ESL 3.12:** Show that the ridge regression estimate can be obtained by ordinary least squares regression on an augmented data set. We augment the centered matrix X with p additional rows $\sqrt{\lambda}I_{p\times p}$, and augment y with p zeros. Comment briefly on how we can think of Ridge shrinkage as adding more "neutral" observations with 0 response.

4. What would be a corresponding case for the Lasso penalty, where the shrinkage can be accomplished by adding data and solving the same fitting problem?
   Hint: Think beyond squared error loss.