Class Exercises. Set 2. Linear Regression

1. We measure the height of a rocket at fixed time intervals as the data pair (s, t). The time t is the time in seconds and s is in meters. The observations are: for times 5, 10, 15, 20, 25 and 30, with heights 722, 1073, 1178, 1177, 781, 102 respectively. Suppose a model $s = a + bt + ct^2$.
   a) Compute the values of a, b, and c with linear regression (without regularization).
   b) Compute the values of a, b, and c with linear ridge regression, setting lambda = 0.5.
   c) What's the average squared error for both models in the 6 points?
   d) *An advantage of Ridge Regression is that a unique solution always exists since $(X^tX + \lambda I)$ is always invertible. Remind that to be invertible, a matrix needs to be full rank. Show that $(X^tX + \lambda I)$ is full rank by characterizing its p eigenvalues in terms of the singular values of X^tX and λI.

2. Consider the following data

x_1	x_2	$y(= output)$
368	15	1.7
340	16	1.5
665	25	2.8
954	40	5
331	15	1.3

   a) What's the regression solution for $f(y) = w_1x_1 + w_2x_2$?

   b) Trying to improve the fitting, we collect another feature x_3:

x_1	x_2	x_3	$y(= output)$
368	15	383	1.7
340	16	356	1.5
665	25	690	2.8
954	40	994	5
331	15	346	1.3

   What's now the solution for $f(y) = w_1x_1 + w_2x_2 + w_3x_3$? Is it unique? What's the complete set of solutions?

3. In some contexts, it is interesting to introduce different costs per example in the error function: $L(w) = 1/2 \sum_{n=1}^{N} c_n(y_n - w_0 - w^tx_n)^2$. with $x_n, w \in R^d$ and $c_n \in R^+$,$w_0 \in R$. Generalize the Probabilistic Interpretation as given in the classes to motivate the given loss function.

4. **Consider the probabilistic model for linear regression as discussed in the class. Assume the model without the bias. For a generic xtest observation, what's the variance in the output variable $\hat{y}$? a) What's the variance of $\hat{y}$ in the origin, xtest=0? b) What's the variance of $\hat{y}$ when the norm of xtest is very large?