Assignment 1

Kelvin Xu
COMP652

January 25, 2015

Problem 1. (a). See function load_data

Problem 1. (b). With each datacase as a row

$$Xw = Y$$

By doing this, we can see that we are not changing any of the decision boundaries. We are just are only effecting the regression coefficients, and their corresponding intepretation. This approach also has the added advantage that it can help avoid numerical instabilities.

Finally, normalization of this type, and other similar normalization schemes helps make the model invariant to parameterization. For example, the ridge regression classifier will have very different weights for the same data shifted around because the penalty will shrink the weight matrix often at the expense of accuracy.

Problem 1. (c). See function polyRegress

Problem 1. (d). See function cross_valid_regress

Problem 1. (e).

Problem 1. (f).

Problem 2.. Writing out the log-likelihood:

$$L = \prod_{i=1}^{m} P(y_i \mid x_i; h) P(x_i)$$

We are interested in the argmax of this value, we also can take a log which doesn't change the argmax since it is a monotonically increasing function.

$$\arg\max_w L = \arg\max \log l$$

$$l = \sum_{i=1}^{m} \log P(y_i \mid x_i; w) + \log P(x_i)$$

Discarding the $P(x_i)$ term since it does not depend on w.

$$
\begin{aligned}
l &= \sum_{i=1}^{m} \log P(y_i \mid x_i; w) \\
&= \sum_{i=1}^{m} \log \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\Big(-\frac{1}{2} \frac{(y_i - h_w(x_i))^2}{\sigma_i^2} \Big) \right) \\
&= \sum_{i=1}^{m} \log \Big(\frac{1}{\sqrt{2\pi\sigma_i^2}} \Big) - \sum_{i=1}^{m} \Big(\frac{1}{2} \frac{(y_i - h_w(x_i))^2}{\sigma_i^2} \Big)
\end{aligned}
$$

Dropping the first term since it has no dependence on w

$$l = -\sum_{i=1}^{m} \Big(\frac{1}{2} \frac{(y_i - h_w(x_i))^2}{\sigma_i^2} \Big)$$

We can see here that maximizing this quantity is equivalent to minimizing

$$l = \sum_{i=1}^{m} \Big(\frac{1}{2} \frac{(y_i - h_w(x_i))^2}{\sigma_i^2} \Big)$$

Assuming our hypothesis is a linear classifier $h_w(X) = Xw$

$$l = \sum_{i=1}^{m} \Big(\frac{1}{2} \frac{(y_i - x_i \cdot w)^2}{\sigma_i^2} \Big)$$

We can get the following by writing our expression in matrix form, letting σ be a diagonal matrix containing our required values.

$$l = \sum_{i=1}^{m} \Big((y - Xw)^T \sigma (y - Xw) \Big)$$

Finally taking, a derivative,

$$l = \sum_{i=1}^{m} \Big((y - Xw)^T \sigma (y - Xw) \Big)$$

TODO, finish last part of this

Problem 3. (a).

Problem 3. (b).

Problem 3. (c).

Problem 4.. Taking the hypothesis to be the product of those basis functions as provided

Writing out the likelihood function:

$$L = \prod_{i=1}^{m} P(y_i \mid x_i; h) P(x_i)$$

Following the same line of reasoning up to the