Machine Learning: Homework #4

Due on November 20, 2017 at 09:59am

Professor Dr. Stephan Guennemann

Marc Meissner - 03691673

Problem 1

Load the notebook 04_*homework_linear_regression.ipynb* from Piazza. Fill in the missing code and run the notebook. Convert the evaluated notebook to pdf and add it to the printout of your homework.

Solution
PDF added to the end of homework.

Problem 2

Let's assume we have a dataset where each datapoint, (x_i, y_i) is weighted by a scalar factor which we will call t_i. We will assume that $t_i > 0$ for all i. This makes the sum of squares error function look like the following:

$E_{\text{weighted}}(w) = \frac{1}{2}\sum_{i=1}^{N} t_i[w^T\Phi(x_i) - y_i]^2$

Find the equation for the value of w that minimizes this error function. Furthermore, explain how this weighting factor, t_i, can be interpreted in terms of
1) the variance of the noise on the data and
2) data points for which there are exact copies in the dataset.

Solution
Writing the equation above in matrix form and setting the derivative to zero:

$E_{\text{weighted}}(W) = \frac{1}{2}(Y - \Phi W)^T T(Y - \Phi W)$

$0 = \frac{\partial}{\partial w}\frac{1}{2}(Y^T - W^T\Phi^T)T(Y - \Phi W)$
$= \frac{\partial}{\partial w}\frac{1}{2}(Y^T TY - Y^T T\Phi W - W^T\Phi^T TY + W^T\Phi^T T\Phi W)$
$= \frac{\partial}{\partial w}\frac{1}{2}(-2Y^T T\Phi W + W^T\Phi^T T\Phi W)$
$= -Y^T T\Phi + \frac{1}{2}(\Phi^T T\Phi + (\Phi^T T\Phi)^T)W$

$W = (\Phi^T T\Phi)^{-1}\Phi^T TY$

t_i can be interpreted in two ways:

1) The higher t_i, the more trust we put into the data point (x_i, y_i). Therefore, it can be seen as an inverse variance parameter that models the uncertainty of individual data points.
2) Imagine $t_i = 2$ for a certain i. The result of the error function would be the same as if we had the i'th data point twice. Consequently, t_i can be seen as the number of identical copies of the i'th data point.

Problem 3

Show that the following holds: The ridge regression estimates can be obtained by ordinary least squares regression on an augmented dataset: Augment the design matrix $\Phi \in \mathbb{R}^{NxM}$ with M additional rows $\sqrt{\lambda} I_{MxM}$ and augment y with M zeros.

Solution

$E_{\text{normal}}(w) = \frac{1}{2}(\Phi w - y)^T(\Phi w - y)$

With augmented matrices:

$$= \frac{1}{2}\left[\begin{pmatrix} \Phi & & & \\ \hline \sqrt{\lambda} & & & \\ & \sqrt{\lambda} & & \\ & & \dots & \\ & & & \sqrt{\lambda} \end{pmatrix} w - \begin{pmatrix} y \\ \hline 0 \end{pmatrix}\right]^T \left[\begin{pmatrix} \Phi & & & \\ \hline \sqrt{\lambda} & & & \\ & \sqrt{\lambda} & & \\ & & \dots & \\ & & & \sqrt{\lambda} \end{pmatrix} w - \begin{pmatrix} y \\ \hline 0 \end{pmatrix}\right]$$

$$= \frac{1}{2}\left[\begin{pmatrix} \Phi w - y \\ \hline \sqrt{\lambda} w \end{pmatrix}\right]^T \left[\begin{pmatrix} \Phi w - y \\ \hline \sqrt{\lambda} w \end{pmatrix}\right]$$

$$= \frac{1}{2}\left[\left((\Phi w - y)^T | (\sqrt{\lambda} w)^T\right)\right] \left[\begin{pmatrix} \Phi w - y \\ \hline \sqrt{\lambda} w \end{pmatrix}\right]$$

$$= \frac{1}{2}\left[(\Phi w - y)^T(\Phi w - y) + (\sqrt{\lambda} w)^T(\sqrt{\lambda} w)\right]$$

$$= \frac{1}{2}\left[(\Phi w - y)^T(\Phi w - y) + \lambda w^T w\right]$$

$$= E_{\text{ridge}}(W)$$

Problem 4

It turns out that the conjugate prior for the situation when we have an unknown mean and unknown precision is a normal-gamma distribution (See section 2.3.6 in Bishop). This is also true when we have a conditional Gaussian distribution of the linear regression model. This means that if our likelihood is as follows:

$$p(y|\Phi, w, \beta) = \prod_{i=1}^{N} \mathcal{N}(y_i | w^T \Phi(x_i), \beta^{-1})$$

Then the conjugate prior for both w and β is

$$p(w, \beta) = \mathcal{N}(w | m_0, \beta^{-1} S_0)\text{Gamma}(\beta | a_0, b_0)$$

Show that the posterior distribution takes the same form as the prior, i.e.

$$p(w, \beta | D) = \mathcal{N}(w | m_N, \beta^{-1} S_N)\text{Gamma}(\beta | a_N, b_N)$$

Also be sure to give the expressions for m_N , S_N , a_N , and b_N.

Problem 4 continued on next page...

Solution

Solution