07: Multivariate In-Class Proofs

Jacob Cluff

March 2019

1

Derive the Normal Equation Solution for fitting an L2 Regularized Multivariate Linear Regression model.

$$
\begin{aligned}
J &= \frac{1}{2N} \sum_{i=1}^{N} \sum_{k=1}^{K} \left(y_{ik} - \hat{y}_{ik}\right)^2 + \frac{\lambda_2}{2N} \sum_{j=1}^{P} \sum_{k=1}^{K} w_{jk}^2 \\
&= \frac{1}{2N} \left\| Y - \hat{Y} \right\|_F^2 + \frac{\lambda_2}{2N} \|W\|_F^2 \\
&= \frac{1}{2N} \operatorname{Tr}\left\{ \left(Y - \hat{Y}\right)^T \left(Y - \hat{Y}\right) \right\} + \frac{\lambda_2}{2N} \operatorname{Tr}\{W^T W\} \\
&= \frac{1}{2N} \operatorname{Tr}\left\{ Y^T Y - Y^T \hat{Y} - \hat{Y}^T Y + \hat{Y}^T \hat{Y} \right\} + \frac{\lambda_2}{2N} \operatorname{Tr}\{W^T W\} \\
&= \frac{1}{2N} \operatorname{Tr}\left\{ Y^T Y - 2Y^T \hat{Y} + \hat{Y}^T \hat{Y} \right\} + \frac{\lambda_2}{2N} \operatorname{Tr}\{W^T W\} \\
&= \frac{1}{2N} \operatorname{Tr}\left\{ Y^T Y - 2Y^T \Phi W + (\Phi W)^T (\Phi W) \right\} + \frac{\lambda_2}{2N} \operatorname{Tr}\{W^T W\} \\
&= \frac{1}{2N} \operatorname{Tr}\{ Y^T Y - 2Y^T \Phi W + W^T \Phi^T \Phi W \} + \frac{\lambda_2}{2N} \operatorname{Tr}\{W^T W\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J}{\partial W} = 0 &= \frac{1}{2N} \left[-2Y^T \Phi + 2W^T \Phi^T \Phi + 2\lambda_2 W^T \right] \\
Y^T \Phi &= W^T \Phi^T \Phi + \lambda_2 W^T \\
\Phi^T Y &= \Phi^T \Phi W + \lambda_2 W \\
\Phi^T Y &= \left(\Phi^T \Phi + \lambda_2 I\right) W \\
W &= \left(\Phi^T \Phi + \lambda_2 I\right)^{-1} \Phi^T Y
\end{aligned}
$$

2

Prove that L2 Regularization yields a Maximum a Posteriori estimator for the model parameters in the Multivariate case.

$$
\begin{aligned}
F &= \exp\{-J\} \\
&= \exp\left\{-\sum_{i=1}^{N}\sum_{k=1}^{K}(y_{ik}-\hat{y}_{ik})^2 - \lambda_2\sum_{j=1}^{P}\sum_{k=1}^{K}w_{jk}^2\right\} \\
&= \exp\left\{-\sum_{i=1}^{N}\sum_{k=1}^{K}(y_{ik}-\hat{y}_{ik})^2\right\}\exp\left\{-\lambda_2\sum_{j=1}^{P}\sum_{k=1}^{K}w_{jk}^2\right\} \\
&= \prod_{i=1}^{N}\prod_{k=1}^{K}\exp\{-(y_{ik}-\hat{y}_{ik})^2\}\prod_{j=1}^{P}\prod_{k=1}^{K}\exp\{-\lambda_2 w_{jk}^2\} \\
&\propto p(Y|\hat{Y})\;p(W) \\
&\propto p(Y|X,W)\;p(W) \\
&\propto p(W|X,Y)
\end{aligned}
$$

The third line converts the product of the product of the product of the two distributions in terms of the probability of getting Y given $\hat{Y}$ ($p(Y|\hat{Y})$) multiplied by the probability of getting the weights ($p(W)$). The second to last line converts $\hat{Y}$ into XW, and the very last line uses Bayes' theorem to find the probability of getting W given a the data and the target. The Probability $p(Y|\hat{Y})$ forms the product of two Gaussian distributions. The probability $p(W)$ is also the product of two Gaussians with a mean of 0 and a variance of $\frac{1}{2\lambda_2}$.

3

Prove that L1 Regularization yields a Maximum a Posteriori estimator for the model parameters in the Multivariate case.

$$
\begin{aligned}
F &= \exp\{-J\} \\
&= \exp\left\{-\sum_{i=1}^{N}\sum_{k=1}^{K}(y_{ik}-\hat{y}_{ik})^2 - \lambda_1\sum_{j=1}^{P}\sum_{k=1}^{K}|w_{jk}|\right\} \\
&= \exp\left\{-\sum_{i=1}^{N}\sum_{k=1}^{K}(y_{ik}-\hat{y}_{ik})^2\right\}\exp\left\{-\lambda_1\sum_{j=1}^{P}\sum_{k=1}^{K}|w_{jk}|\right\} \\
&= \prod_{i=1}^{N}\prod_{k=1}^{K}\exp\{-(y_{ik}-\hat{y}_{ik})^2\}\prod_{j=1}^{P}\prod_{k=1}^{K}\exp\{-\lambda_2 w_{jk}\} \\
&\propto p(Y|\hat{Y})\ p(W) \\
&\propto p(Y|X,w)\ p(W) \\
&\propto p(w|X,y)
\end{aligned}
$$

This is solved very similar to the previous problem; however, the probability distribution for W is quite different. Instead of being the product of two Gaussian distributions, $p(W)$ is the product of two Laplacian distributions with a mean of 0 and a diversity of $\frac{1}{\lambda_1}$