Comp 6321 - Machine Learning - Assignment 2

Federico O'Reilly Regueiro

October 18th, 2016

Question 1:

1.a Partition the data into training / testing, 90 to 10 and perform and plot $L2$ regularization

Lorem Ipsum

1.b Use the `quadprog` function in Matlab for $L1$ regularization

$$\arg\min_{w} \frac{1}{2}(\boldsymbol{\Phi w} - \boldsymbol{y})^T(\boldsymbol{\Phi w} - \boldsymbol{y}) + \frac{\lambda}{2}\sum_{k=0}^{K-1}|w_k| \tag{1}$$

Which is equivalent to finding:

$$\arg\min_{w} (\boldsymbol{\Phi w} - \boldsymbol{y})^T(\boldsymbol{\Phi w} - \boldsymbol{y}) + \lambda\sum_{k=0}^{K-1}|w_k| \tag{2}$$

And expands to:

$$\arg\min_{w} \boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi w} - 2y^T\boldsymbol{\Phi w} + \boldsymbol{y}^T\boldsymbol{y}^T + \lambda\sum_{k=0}^{K-1}|w_k| \tag{3}$$

And for which we can remove the constant term $\boldsymbol{y}^T\boldsymbol{y}$, yielding:

$$\arg\min_{w} \boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi w} - \boldsymbol{2y}^T\boldsymbol{\Phi w}^T + \lambda\sum_{k=0}^{K-1}|w_k| \tag{4}$$

Matlab's `quadprog(H, f, A, b)` function, gives the optimal $\boldsymbol{x}$ corresponding to the expression $\arg\min_{x} \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Hx} + \boldsymbol{f}^T\boldsymbol{x}$, subject to constraints $\boldsymbol{Ax} \leq \boldsymbol{b}$. We can thus take `H :=` $2\boldsymbol{\Phi}^T\boldsymbol{\Phi}$, then `f :=` $2\boldsymbol{y}^T\boldsymbol{\Phi}$, `A :=`$\lambda\boldsymbol{P}$, where for a system with n variables, $\boldsymbol{P}$ is the matrix with 2^n permutations of $[b_1, b_2, \ldots b_n, 0]$[1], $b \in \{-1, 1\}$

[1] the last zero entry is to avoid the regularization of the intercept term

and lastly `b := ` $c\ \vec{\mathbf{1}}$, where $\vec{\mathbf{1}}$ is an all-one vector of length 2^n that places an upper bound to the expression $\lambda \sum_{k=0}^{K-1} |w_k|$, such that $\sum_{k=0}^{K-1} |w_k| \leq \frac{c}{\lambda}$. If we set $c = 1$, then we parametrize $L1$ regularization simply by adjusting λ accordingly.