Comp 6321 - Machine Learning - Assignment 2

Federico O'Reilly Regueiro

October 18th, 2016

Question 1:

1.a Doodles

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

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 $\arg\min_{x} \boldsymbol{x}^T\boldsymbol{Hx} + \boldsymbol{f}^T\boldsymbol{x}$, subject to constraints $\boldsymbol{Ax} \leq \boldsymbol{b}$. We can thus take `H :=` $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$, 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 $[b_1, b_2, \ldots b_n], b \in \{-1, 1\}$ and lastly `b :=` $c\ \overrightarrow{\boldsymbol{1}}$ places an upper bound to the expression $\lambda\sum_{k=0}^{K-1}|w_k|$, such that $\sum_{k=0}^{K-1}|w_k| \leq \frac{c}{\lambda}$ TODO develop relationship between lambda, c and whatnot, too tired to think now...

This here, on the other hand, is something else...

$$\begin{aligned} &\min_{w} J_D(\mathbf{w}) = \min_{w}(\boldsymbol{\Phi w} - \boldsymbol{y})^T(\boldsymbol{\Phi w} - \boldsymbol{y}) \\ &s.t. \sum_{i=1}^{n}|w_i| \leq \eta \end{aligned} \tag{5}$$