Tsinghua-Berkeley Shenzhen Institute
Learning from Data
Fall 2019

Written Assignment 5

Issued: Saturday 14th December, 2019 **Due:** Friday 3rd January, 2020

5.1. (2 points) Using the technique of Lagrange multipliers, show that minimization of the regularized error function

$$\frac{1}{2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i\right)^2 + \frac{\lambda}{2}\sum_{i=1}^{n}|w_i|^q$$

is equivalent to minimizing the unregularized sum-of-squares error

$$\frac{1}{2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i\right)^2$$

subject to the constraint

$$\sum_{j=1}^{n}|w_j|^q \le \eta$$

5.2. (2 points) (MAP Estimation) We mentioned that when θ is an isotropic Laplace distribution, MAP corresponds to LASSO (L1-regularization). Now you are maximizing the likelyhood function $\prod_{i=1}^{n} p(x_i|\theta)$ with prior distribution $p(\theta)$.

$$p(\theta) = \frac{1}{2}\lambda \exp\left(-\lambda|\theta|\right) \qquad \lambda > 0$$

Please prove that it is equivalent to maximizing

$$\log \prod_{i=1}^{n} p(x_i|\theta) - \lambda|\theta|$$

5.3. (2 points) (Mean Square Error) We mentioned Bias-Variance Tradeoff in class. We define the MSE of $\hat{X}$, an estimator of X as $\mathrm{MSE}(\hat{X}) \triangleq \mathbb{E}[(\hat{X}-X)^2]$. The variance of $\hat{X}$ is defined as $\mathrm{Var}(\hat{X}) \triangleq \mathbb{E}[(\hat{X}-\mathbb{E}[\hat{X}])^2]$ and the bias is defined as $\mathrm{Bias}(\hat{X}) \triangleq \mathbb{E}[\hat{X}]-X$.

(a) Please prove that

$$\mathrm{MSE}(\hat{X}) = \mathrm{Var}(\hat{X}) + (\mathrm{Bias}(\hat{X}))^2$$

(b) Our data are added with an independent Gaussian noise, say, $X + N$, where $\mathbb{E}[N] = 0$ and $\mathbb{E}[N^2] = \sigma^2$ and the estimator is $\hat{X}$. We define the empirical MSE as $\mathbb{E}[(\hat{X} - X - N)^2]$. Please prove that

$$\mathbb{E}[(\hat{X} - X - N)^2] = \mathrm{MSE}(\hat{X}) + \sigma^2$$

The equation tells us that the empirical error is a good estimation of the true error. Thus, we can minimize the empirical error in order to properly minimize the true error.

5.4. (4 points) (VC Dimension) Given some finite domain set, $\mathcal{X}$, and a number $k \leq |\mathcal{X}|$, please figure out the VC-dimension of each of the following classes:

(a) (2 points) $\mathcal{H}_k^{\mathcal{X}} = \left\{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x) = 1\}| = k\right\}$. That is, the set of all functions that assign the value 1 to exactly k elements of $\mathcal{X}$.

(b) (2 points) $\mathcal{H}_{\leq k}^{\mathcal{X}} = \left\{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x) = 1\}| \leq k \text{ or } |\{x : h(x) = 0\}| \leq k\right\}$

5.5. (Bonus) Please write down some brief thoughts or summary of the WODS Keynote Speech you listen. Words more than half page are required.