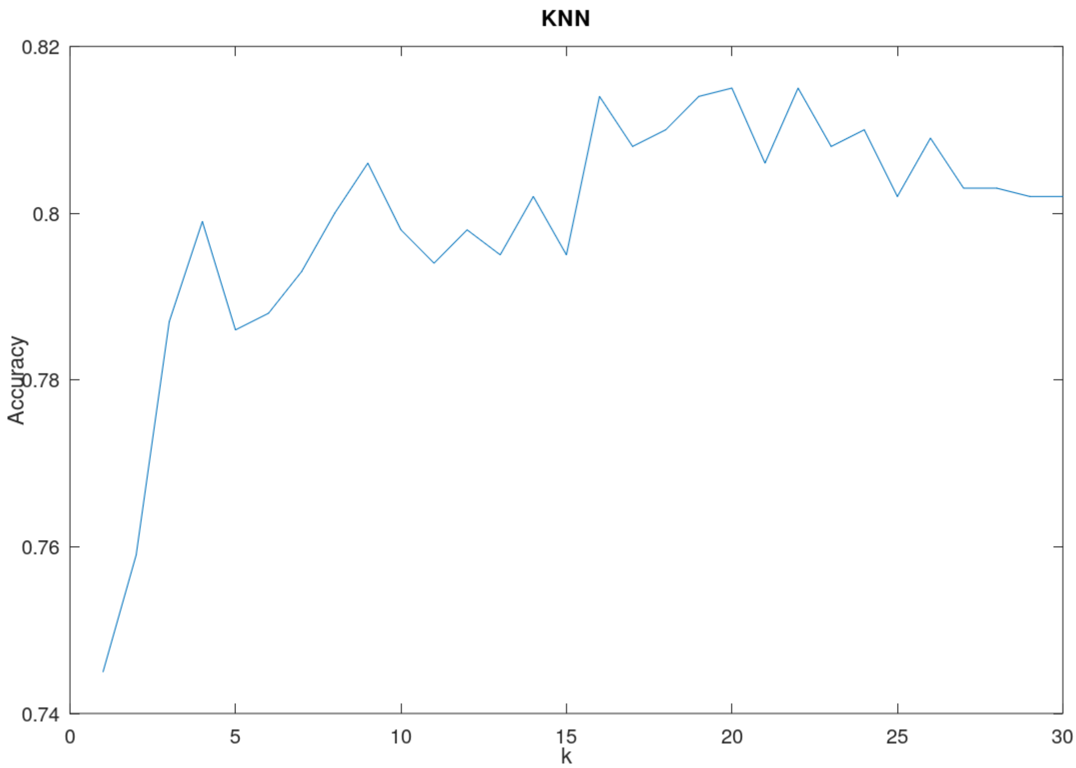
CS 480/680 Spring 2019 - Assignment 1 Solutions

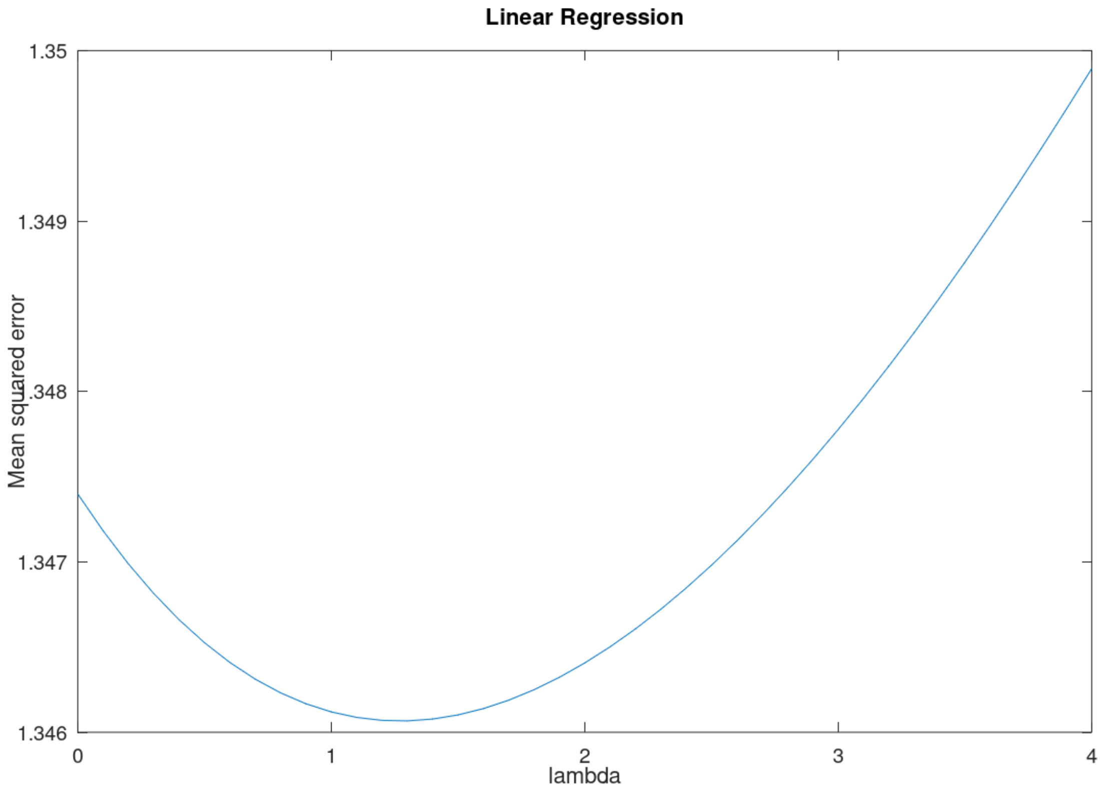
June 6, 2019

Question 1



The best k is 20 with cross validation accuracy 81.5%. The test accuracy for k=20 is 76.4%. The result is based on Euclidean distance. Any solution with a similar graph was accepted as a correct solution.

Question 2



The best λ is 1.3 with cross validation error of 1.346. The test mean squared error for $\lambda = 1.3$ is 1.436.

Question 3

Part (a)

Write the objective in matrix form

$$\begin{aligned} L(\mathbf{w}, b) &= \sum_{n=1}^{N} r_n \big(y_n - \mathbf{w}^T \mathbf{x}_n + b\big)^2 \\ &= (\mathbf{y} - \mathbf{v}^T \bar{\mathbf{X}})\mathbf{R}(\mathbf{y} - \mathbf{v}^T \bar{\mathbf{X}})^T \\ &= \mathbf{y}\mathbf{R}\mathbf{y}^T - 2\mathbf{y}\mathbf{R}\bar{\mathbf{X}}^T \mathbf{v} + \mathbf{v}^T \bar{\mathbf{X}}\mathbf{R}\bar{\mathbf{X}}^T \mathbf{v} \end{aligned}$$

where

$$\mathbf{v} = \begin{bmatrix} -b \\ \mathbf{w} \end{bmatrix}$$

$$\mathbf{y} = \big[y_1, ..., y_N\big]$$

$$\bar{\mathbf{X}} = \begin{bmatrix} 1..........1 \\ \mathbf{x}_1, ..., \mathbf{x}_N \end{bmatrix}$$

$\mathbf{R}$ is a diagonal matrix with entries $diag(r_1, r_2, ...r_N)$
Assuming the weights r_n are positive, the objective is a convex quadratic function of $\mathbf{w}$ and b. So, the minimum occurs at the point of zero gradient :

$$0 = \nabla L(\mathbf{v}) = -2\bar{\mathbf{X}}\mathbf{R}\mathbf{y}^T + 2\bar{\mathbf{X}}\mathbf{R}\bar{\mathbf{X}}^T \mathbf{v}$$

Solve for $\mathbf{v}$ and we get

$$\mathbf{v} = (\bar{\mathbf{X}}\mathbf{R}\bar{\mathbf{X}}^T)^{-1}\bar{\mathbf{X}}\mathbf{R}\mathbf{y}^T$$

Part (b)

Suppose $y_n = \mathbf{w}^T \mathbf{x}_n - b + \epsilon_n$ where ϵ_n is a zero-mean Gaussian random variable with variance σ_n^2. Thus, the conditional density of y_n given $\mathbf{x}_n$ is Gaussian with mean $\mathbf{w}^T \mathbf{x}_n - b$ and variance σ_i^2. The negative log-likelihood is

$$\begin{aligned} -l(\mathbf{w}, b) &= -\log Pr(y_1, ..y_N | \mathbf{x}_1, ..., \mathbf{x}_N) \\ &= -\log \prod_{n=1}^{N} Pr(y_n | \mathbf{x}_n) \\ &= -\sum_{n=1}^{N} \log Pr(y_n | \mathbf{x}_n) \\ &= \sum_{n=1}^{N} \log(\sqrt[2]{2\pi\sigma_n^2}) + \sum_{n=1}^{N} \frac{(y_n - \mathbf{w}^T \mathbf{x}_n + b)^2}{2\sigma_n^2} \end{aligned}$$

The first term is constant with respect to $\mathbf{w}$ and b. So minimizing the negative log-likelihood is equivalent to minimizing only the second term, which is the same as our objective function with $r_n = \frac{1}{2\sigma_n^2}$. The variance of measurement n in this model is thus inversely proportional to the weight r_n : $\sigma_n^2 \propto \frac{1}{r_n}$