# CS 698: Assignment 3

Ronghao Yang
ID:20511820
Session: 4:00pm-5:20pm

November 1, 2017

## 1 Exercise 1

### 1.1 Question 1.1

$$\text{let's define } p(x_i) = \frac{1}{1+e^{-y_i w^T x_i}}$$

$$\text{let } g_k = \frac{\partial}{\partial w_k} f(w) = \frac{1}{n^+} \sum_{i:y_i=1} \frac{-e^{-y_i w^T x_i} y_i x_i^k}{1+e^{-y_i w^T x_i}} + \frac{1}{n^-} \sum_{j:y_j=-1} \frac{-e^{-y_j w^T x_j} y_j x_j^k}{1+e^{-y_j w^T x_j}} + 2\lambda w_k$$

$$\text{Therefore, } g(k) = \frac{1}{n^+} \sum_{i:y_i=1} -p(x_i)\left(\frac{1}{p(x_i)} - 1\right) y_i x_i^k + \frac{1}{n^-} \sum_{i:y_j=-1} -p(x_j)\left(\frac{1}{p(x_j)} - 1\right) y_j x_j^k + 2\lambda w_k$$

$$\nabla f(w) = \begin{bmatrix} g_1 \\ g_2 \\ \ldots \\ g_d \end{bmatrix}$$

$$\text{let } s_{ks} = \frac{\partial}{\partial w_s} g_k$$

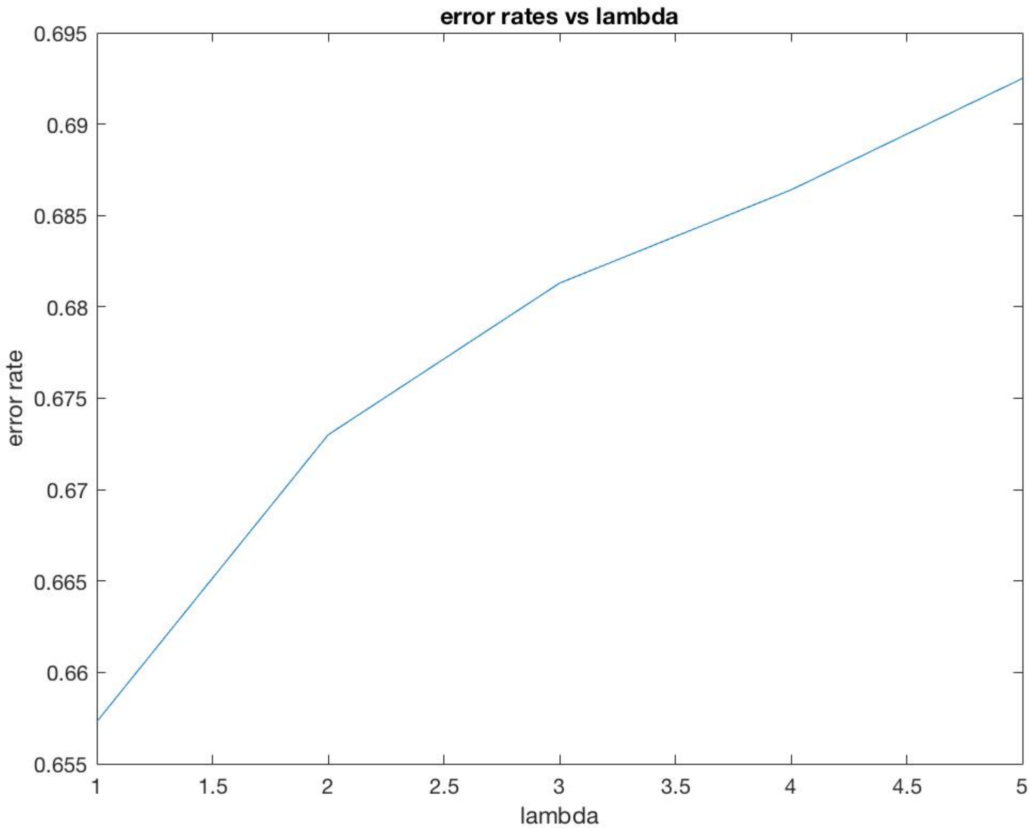
$$\text{Then } s_{ks} = \frac{1}{n^+} \sum_{i:y_i=1} \frac{e^{-y_i w^T x_i} y_i^2 x_i^k x_i^s}{(1+e^{-y_i w^T x_i})^2} + \frac{1}{n^-} \sum_{j:y_j=-1} \frac{e^{-y_j w^T x_j} y_j^2 x_j^k x_j^s}{(1+e^{-y_j w^T x_j})^2} + 2\lambda\epsilon, \text{ where } \epsilon = 1 \text{ if } s = k, \text{ else } \epsilon = 0$$

$$\text{Therefore, } s_{ks} = \frac{1}{n^+} \sum_{i:y_i=1} (p(x_i) - p^2(x_i)) x_i^k x_i^s + \frac{1}{n^-} \sum_{j:y_j=-1} (p(x_j) - p^2(x_j)) x_j^k x_j^s + 2\lambda\epsilon$$

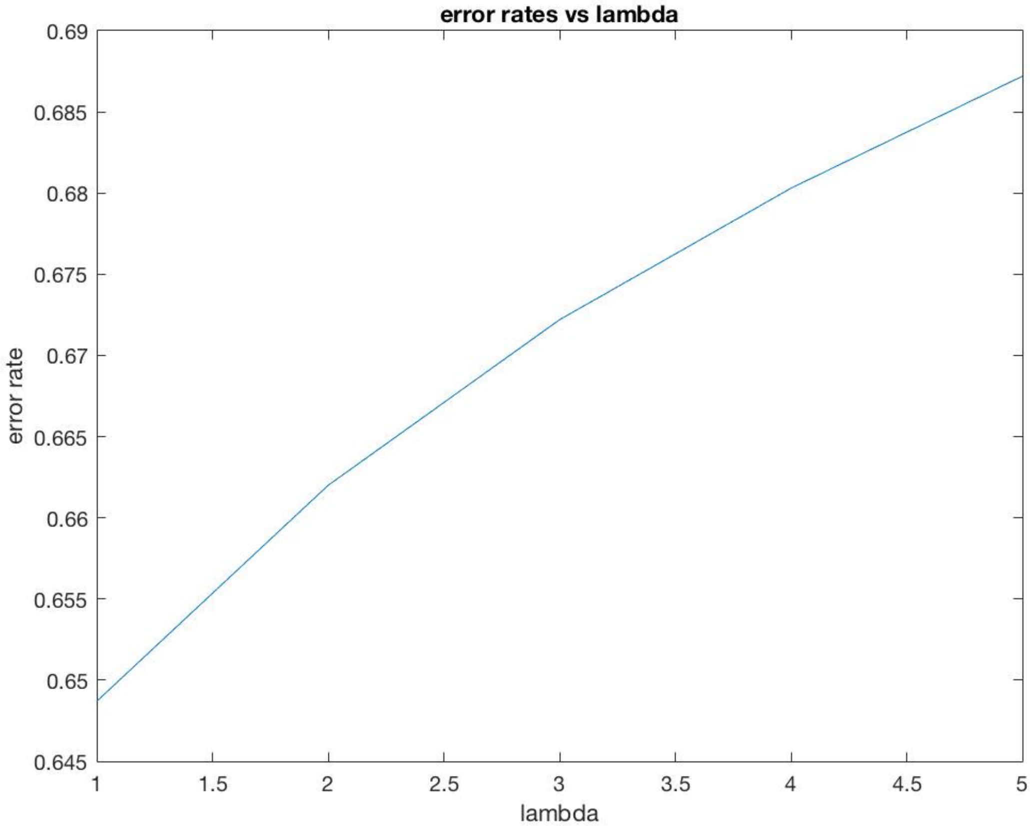
$$\text{where } \epsilon = 1 \text{ if } s = k, \text{ else } \epsilon = 0$$

$$\nabla^2 f(w) = \begin{bmatrix} s_{11} & s_{12} & \ldots & s_{1d} \\ s_{21} & s_{22} & \ldots & s_{2d} \\ \ldots & & & \\ s_{d1} & s_{d2} & \ldots & s_{dd} \end{bmatrix}$$

## 1.2 Question 1.2



## 1.3 Question 1.3




## 1.4 Question 1.4

For kernelized logistic regression, each $x$ is transformed to $\phi(x)$, and the weight is defined in terms of support vectors

$$w = \sum_i \alpha_i \phi(x_i)$$

Therefore, the sigmoid function is defined as

$$p = \frac{1}{1+e^{-y_i \sum_i \alpha_i \phi(x_i)\phi(x)}} = \frac{1}{1+e^{-y_i \sum_i \alpha_i K(x,x_i)}}$$

The the loss function of kernel logistic regression becomes:

$$\min_w \frac{1}{n^+} \sum_{i:y_i=1} log(1 + exp(-y_i \sum_l \alpha_l K(x, x_i))) + \frac{1}{n^-} \sum_{j:y_j=-1} log(1 + exp(-y_j \sum_l \alpha_l K(x, x_j))) + \lambda w^t w$$

To calculate the Gradient, we do the following:

$$\nabla f(\alpha) = \frac{1}{n^+}\sum_{i:y_i=1} -p(x_i)(\frac{1}{p(x_i)} - 1)y_i K(x, x_i) + \frac{1}{n^-}\sum_{i:y_j=-1} -p(x_j)(\frac{1}{p(x_j)} - 1)y_j K(x, x_j) + 2\lambda K\alpha$$

To calculate the Hessian, we do the following:

$$\text{let } s_{ks} = \frac{\partial}{\partial \alpha_s} g_k$$

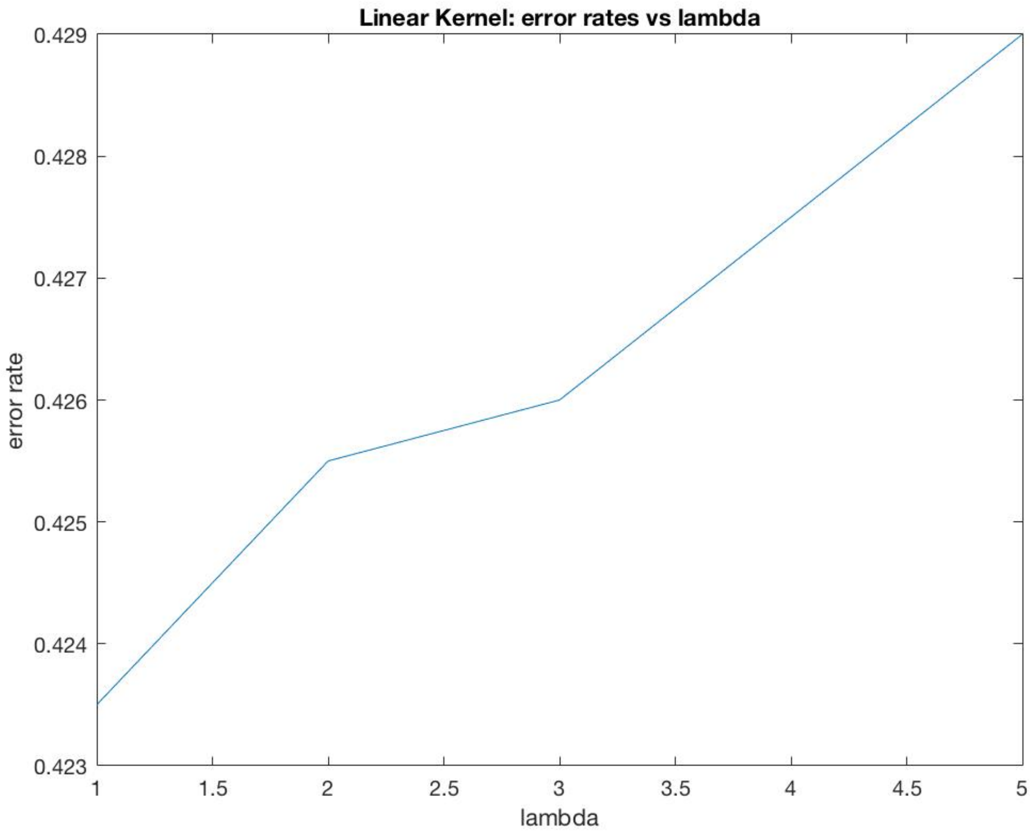
$$\text{Then } s_{ks} = \frac{1}{n^+}\sum_{i:y_i=1} \frac{e^{-y_i \sum_l \alpha_l k(x,x_i)} K(x_k,x_i) K(x_s,x_i)}{(1+e^{-y_i \sum_l \alpha_l K(x,x_i)})^2} + \frac{1}{n^-}\sum_{j:y_j=-1} \frac{e^{-y_i \sum_l \alpha_l k(x,x_j)} K(x_k,x_j) K(x_s,x_j)}{(1+e^{-y_j \sum_l \alpha_l K(x,x_j)})^2} + 2\lambda K(x_k, x_s)$$

Therefore

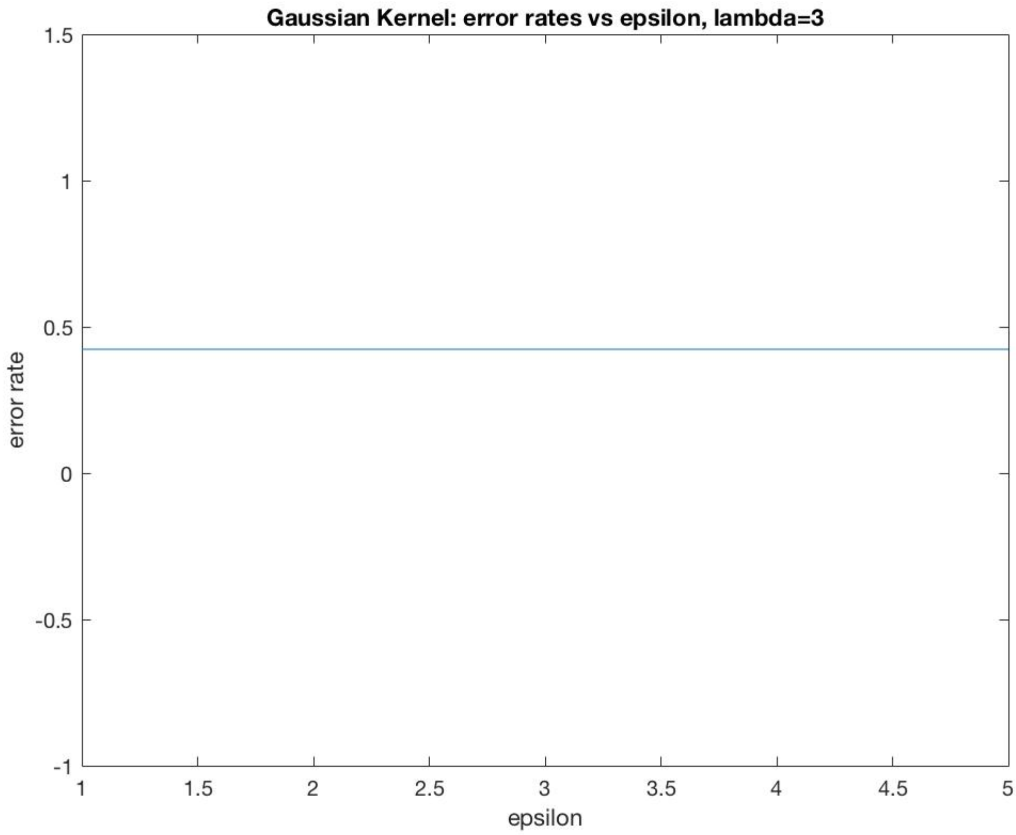
$$s_{ks} = \frac{1}{n^+}\sum_{i:y_i=1}(p(x_i) - p^2(x_i))K(x_k, x_i)K(x_s, x_i) + \frac{1}{n^-}\sum_{j:y_j=-1}(p(x_j) - p^2(x_j))K(x_k, x_j)K(x_s, x_j) + 2\lambda K(x_k, x_s)$$

$$\nabla^2 f(\alpha) = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1d} \\ s_{21} & s_{22} & \dots & s_{2d} \\ \dots & & & \\ s_{d1} & s_{d2} & \dots & s_{dd} \end{bmatrix}$$

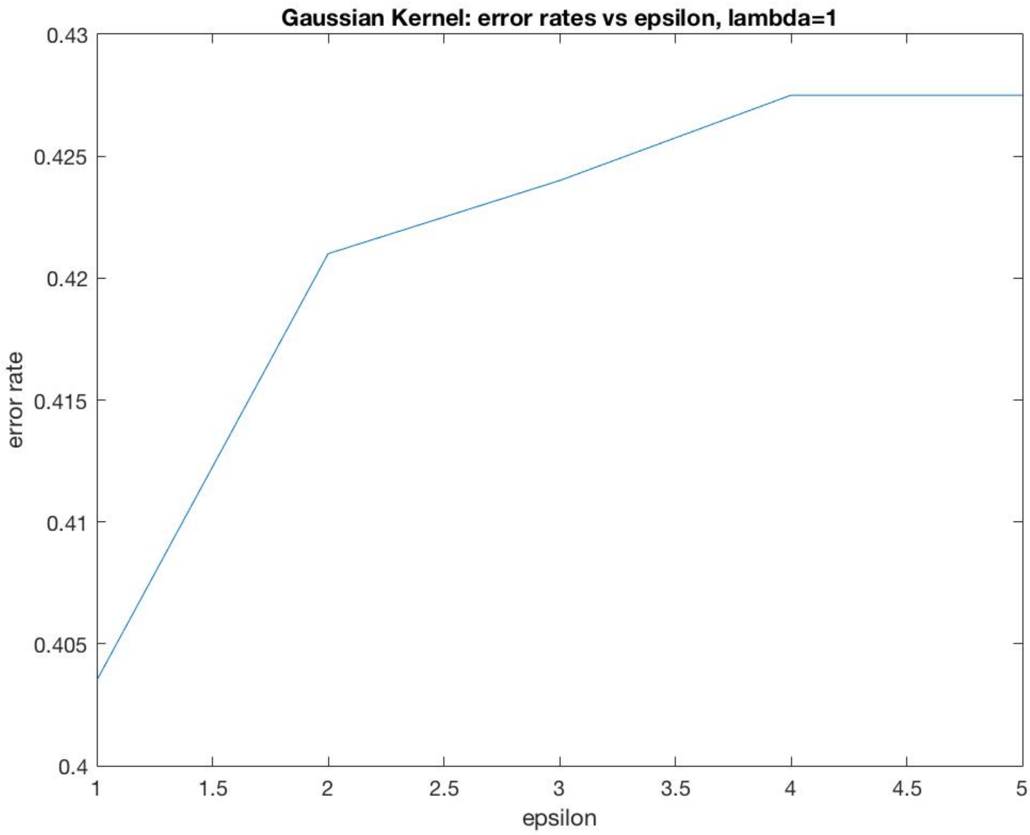
For linear kernel, we have:



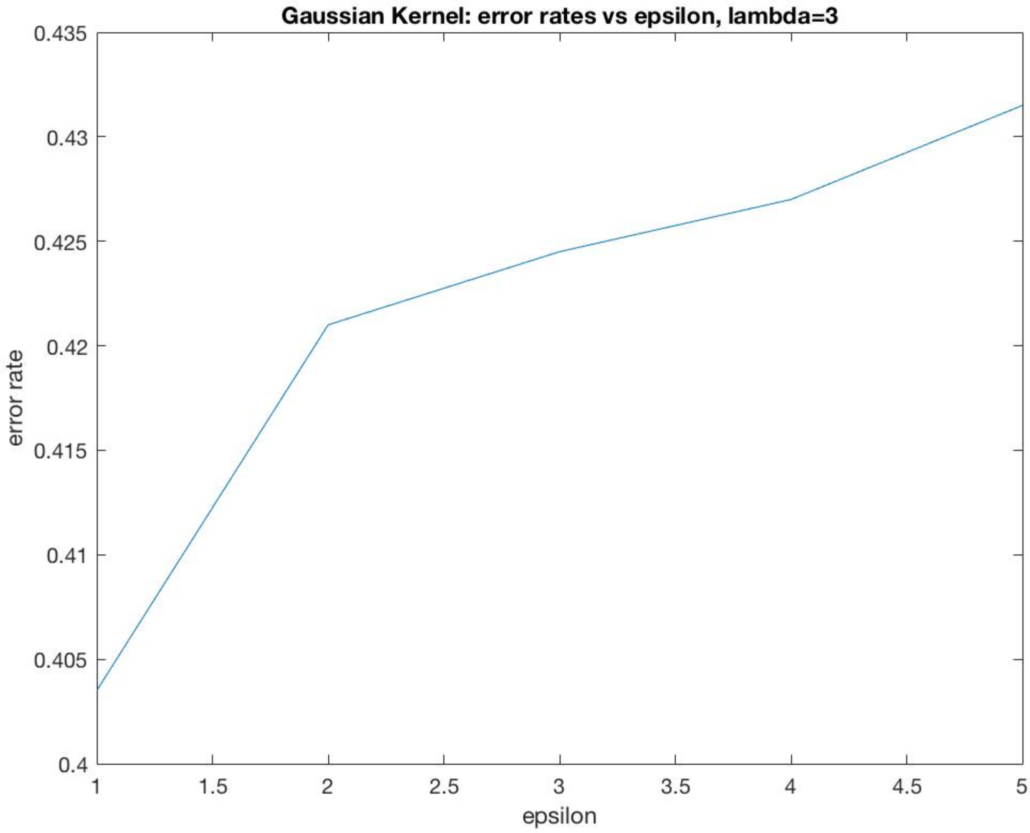

For polynomial kernel, we have(The error rate maintains 42.35% for all $\lambda$ values from 1 to 5):



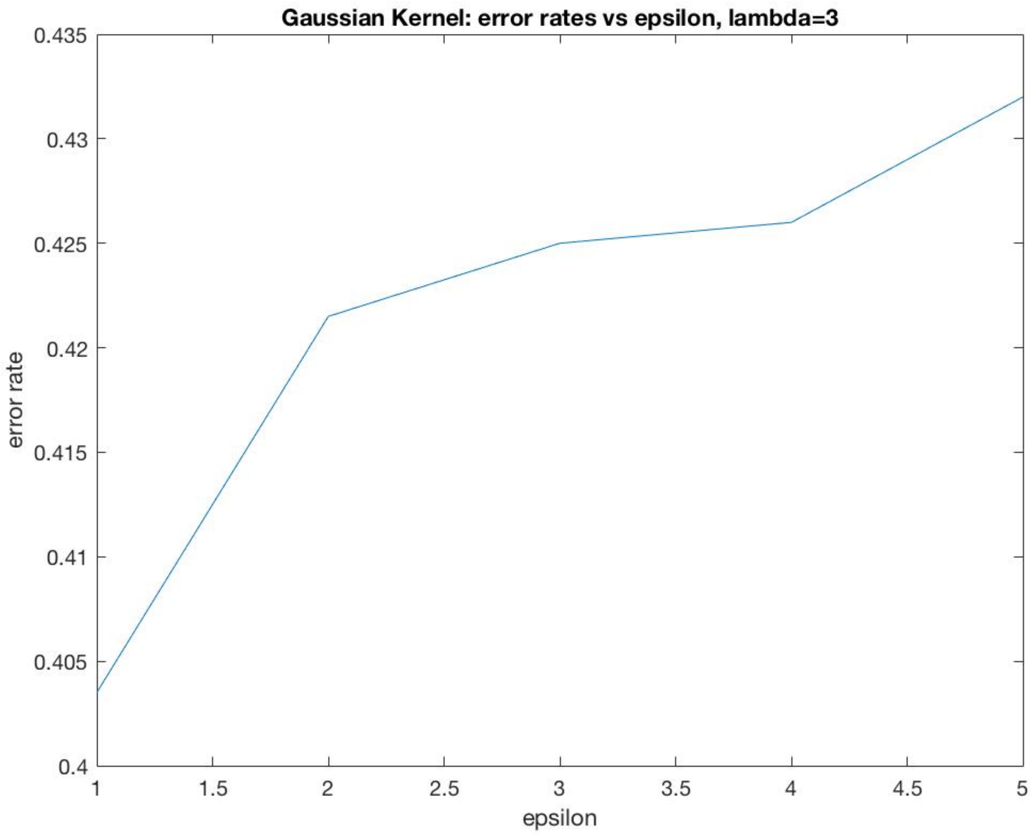

For Gaussian kernel, when $\lambda = 1$, we have:




For Gaussian kernel, when $\lambda = 3$, we have:




For Gaussian kernel, when $\lambda = 5$, we have:




# 2 Exercise 2

## 2.1 Question 2.1

## 2.2 Question 2.2

## 2.3 Question 2.3

# 3 Exercise 3

## 3.1 Question 3.1

## 3.2 Question 3.2

## 3.3 Question 3.3