STAD80: Assignment 5

Vladislav Trukhin

Due: March 17th, 2022

Contents

Question 1

Question 1.1

```
source("/Users/vladislavtrukhin/Downloads/SpamAssassin/functions.R")
```

```
feature <- function(pos, neg) {
  pos_gray <- rgb2gray(pos)
  neg_gray <- rgb2gray(neg)

  neg_crop <- crop.r(neg_gray, 160, 96)

  pos_grad <- grad(pos_gray, 128, 64, FALSE)
  neg_grad <- grad(neg_crop, 128, 64, FALSE)

  pos_fet <- hog(pos_grad[[1]], pos_grad[[2]], 4, 4, 6)
  neg_fet <- hog(neg_grad[[1]], neg_grad[[2]], 4, 4, 6)

  return(list(pos_fet, neg_fet))
}
```

```
fet_data <- c()
pos_data <- c()
for (i in 1:500) {
  pos <- readPNG(
```

```
    sprintf("/Users/vladislavtrukhin/Downloads/A4_datasets/pngdata/pos/%d.png", i))
  neg <- readPNG(
    sprintf("/Users/vladislavtrukhin/Downloads/A4_datasets/pngdata/neg/%d.png", i))
  fet <- feature(pos, neg)
  fet_data <- rbind(fet_data, fet[[1]], fet[[2]])
  pos_data <- rbind(pos_data, 1, 0)
}
```

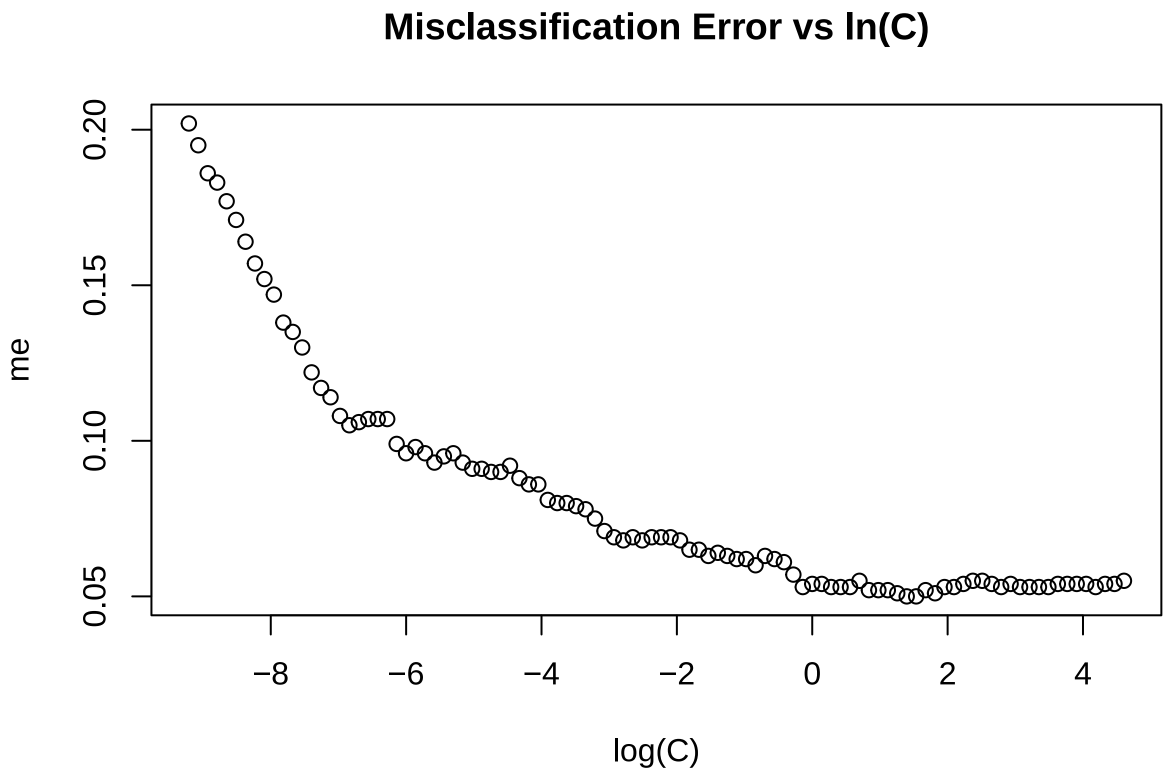
```
C <- exp(seq(log(10^-4), log(10^2), length.out=100))
cve <- c()
me <- c()
for (i in 1:100){
  svm <- ksvm(fet_data, pos_data, type="C-svc", kernel="vanilladot", cross=5, C=C[i])
  cve <- cbind(cve, cross(svm))
  me <- cbind(me, error(svm))
}
```

```
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
```

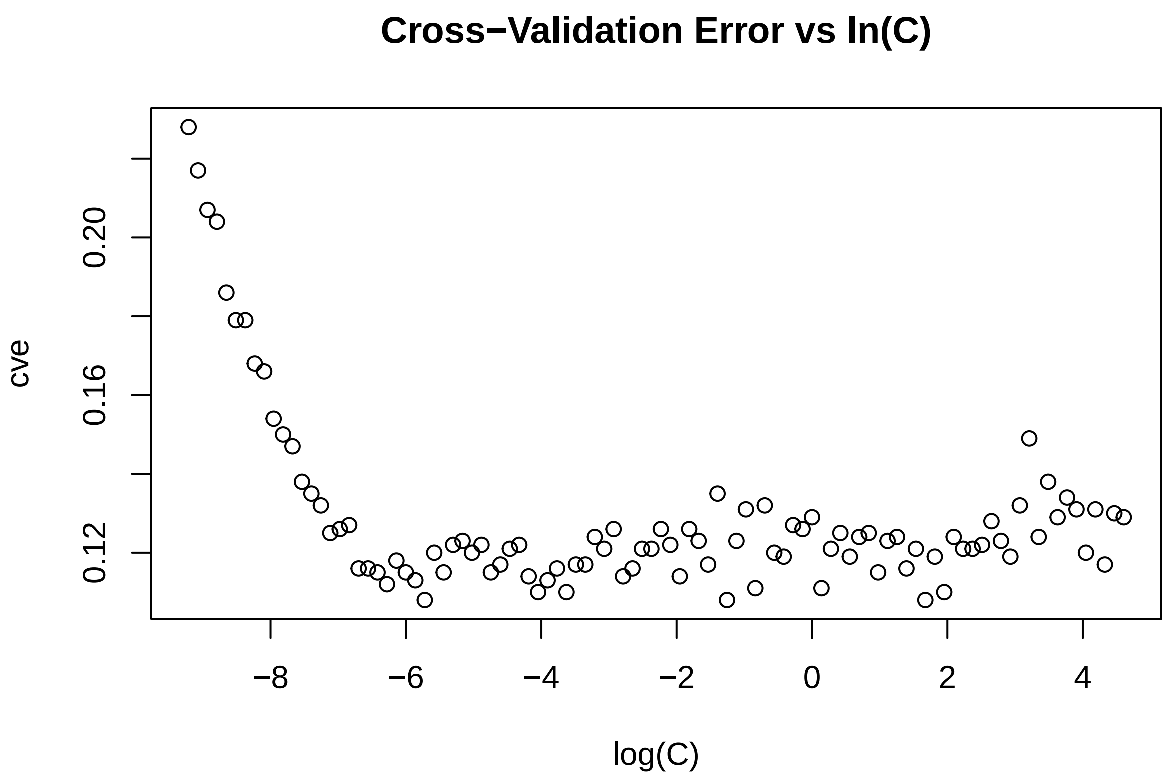
```
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
```

```
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
##  Setting default kernel parameters
```

```
plot(log(C), me)
title("Misclassification Error vs ln(C)")
```




```
plot(log(C), cve)
title("Cross-Validation Error vs ln(C)")
```




```
C[which.min(me)] # Optimal C that yields lowest misclassification error
```

```
## [1] 4.037017
```

The cross validation error decreases as C increases to 10^-5 and increases past 10^-5.

```
cv <- cv.glmnet(fet_data, pos_data, family="binomial", type.measure="class")
min(cv$cvm)
```

```
## [1] 0.114
```

```
min(cve)
```

```
## [1] 0.108
```

The lowest cross validation of SVM is lower than the lowest cross validation of logistic regression, however not significantly.

Question 2

Question 2.1

$\Sigma_{i=1}^n \log p(\mathbf{x_i}, y_i)$

$= \Sigma_{i=1:y_i=1}^n \log p(y_i = 1)p(\mathbf{x_i}|y_i = 1) + \Sigma_{i=1:y_i=2}^n \log p(y_i = 2)p(\mathbf{x_i}|y_i = 2)$

$= \Sigma_{i=1:y_i=1}^n \log \eta \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} exp(\frac{-(\mathbf{x_i}-\mu_1)^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_1)}{2}) + \Sigma_{i=1:y_i=2}^n \log(1-\eta)\frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} exp(\frac{-(\mathbf{x_i}-\mu_2)^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_2)}{2})$

$= \Sigma_{i=1:y_i=1}^n \log \eta - d/2 \log(2\pi) - 1/2 \log|\mathbf{\Sigma}| + \frac{-(\mathbf{x_i}-\mu_1)^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_1)}{2} + \Sigma_{i=1:y_i=2}^n \log(1-\eta) - d/2\log(2\pi) - 1/2 \log|\mathbf{\Sigma}| + \frac{-(\mathbf{x_i}-\mu_2)^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_2)}{2}$

$= n_1 \log \eta - dn/2 \log(2\pi) + n/2 \log|\mathbf{\Sigma}|^{-1} + \Sigma_{i=1:y_i=1}^n \frac{-(\mathbf{x_i}-\mu_1)^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_1)}{2} + n_2 \log(1-\eta) + \Sigma_{i=1:y_i=2}^n \frac{-(\mathbf{x_i}-\mu_2)^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_2)}{2}$

$= n_1 \log \eta + n_2 \log(1-\eta) - dn/2 \log(2\pi) + n/2 \log |\mathbf{\Sigma^{-1}}| + \Sigma_{i=1:y_i=1}^n \frac{-(\mathbf{x_i}-\mu_\mathbf{1})^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_\mathbf{1})}{2} + \Sigma_{i=1:y_i=2}^n \frac{-(\mathbf{x_i}-\mu_\mathbf{2})^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_\mathbf{2})}{2}$

Question 2.2

MLE $\hat{\eta}$

$\frac{\partial}{\partial \eta} \Sigma_{i=1}^n \log p(\mathbf{x_i}, y_i)$

$= n_1 \frac{\partial}{\partial \eta} \log \eta + n_2 \frac{\partial}{\partial \eta} \log(1-\eta)$

$= \frac{n_1}{\eta} - \frac{n_2}{1-\eta}$

$\Rightarrow \frac{n_1}{\hat{\eta}} = \frac{n_2}{1-\hat{\eta}}$

$\Rightarrow n_1(1-\hat{\eta}) = n_2 \hat{\eta}$

$\Rightarrow n_1 = n_2 \hat{\eta} + n_1 \hat{\eta}$

$\Rightarrow \frac{n_1}{n_1+n_2} = \hat{\eta}$

MLE $\hat{\mu}_\mathbf{1}$

$\frac{\partial}{\partial \mu_1} \Sigma_{i=1}^n \log p(\mathbf{x_i}, y_i)$

$= \frac{\partial}{\partial \mu_1} \Sigma_{i=1:y_i=1}^n \frac{-(\mathbf{x_i}-\mu_\mathbf{1})^\top \mathbf{\Sigma}^{-1}(\mathbf{x_i}-\mu_\mathbf{1})}{2}$

$= -1/2 \Sigma_{i=1:y_i=1}^n \frac{\partial}{\partial \mu_1} (\mathbf{x_i} - \mu_\mathbf{1})^\top \mathbf{\Sigma}^{-1} (\mathbf{x_i} - \mu_\mathbf{1})$

$= \Sigma_{i=1:y_i=1}^n (\mathbf{x_i} - \mu_\mathbf{1})^\top \mathbf{\Sigma}^{-1}$

$\Rightarrow \Sigma_{i=1:y_i=1}^n (\mathbf{x_i} - \hat{\mu}_\mathbf{1}) = \mathbf{0}$

$\Rightarrow \frac{\Sigma_{i=1:y_i=1}^n \mathbf{x_i}}{n_1} = \hat{\mu}_\mathbf{1}$

MLE $\hat{\mu}_\mathbf{2}$

Similar case follows as $\hat{\mu}_\mathbf{1}$, $\frac{\Sigma_{i=1:y_i=2}^n \mathbf{x_i}}{n_2} = \hat{\mu}_\mathbf{2}$

Question 2.3

$\frac{\partial}{\partial \mathbf{\Sigma}^{-1}} \Sigma_{i=1}^n \log p(\mathbf{x_i}, y_i)$

$= n/2 \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} \log |\mathbf{\Sigma^{-1}}| - 1/2 \Sigma_{i=1:y_i=1}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} (\mathbf{x_i} - \mu_\mathbf{1})^\top \mathbf{\Sigma}^{-1} (\mathbf{x_i} - \mu_\mathbf{1}) - 1/2 \Sigma_{i=1:y_i=2}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} (\mathbf{x_i} - \mu_\mathbf{2})^\top \mathbf{\Sigma}^{-1} (\mathbf{x_i} - \mu_\mathbf{2})$

$= n/2\Sigma - 1/2 \Sigma_{i=1:y_i=1}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} trace((\mathbf{x_i} - \mu_\mathbf{1})^\top \mathbf{\Sigma}^{-1} (\mathbf{x_i} - \mu_\mathbf{1})) - 1/2 \Sigma_{i=1:y_i=2}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} trace((\mathbf{x_i} - \mu_\mathbf{2})^\top \mathbf{\Sigma}^{-1} (\mathbf{x_i} - \mu_\mathbf{2}))$

$= n/2\Sigma - 1/2 \Sigma_{i=1:y_i=1}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} trace(\mathbf{\Sigma}^{-1} (\mathbf{x_i} - \mu_\mathbf{1})(\mathbf{x_i} - \mu_\mathbf{1})^\top) - 1/2 \Sigma_{i=1:y_i=2}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} trace((\mathbf{\Sigma}^{-1} (\mathbf{x_i} - \mu_\mathbf{2})(\mathbf{x_i} - \mu_\mathbf{2})^\top)$

$= n/2\Sigma - 1/2 \Sigma_{i=1:y_i=1}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} trace((\mathbf{x_i} - \mu_\mathbf{1})(\mathbf{x_i} - \mu_\mathbf{1})^\top \mathbf{\Sigma}^{-1}) - 1/2 \Sigma_{i=1:y_i=2}^n \frac{\partial}{\partial \mathbf{\Sigma}^{-1}} trace(((\mathbf{x_i} - \mu_\mathbf{2})(\mathbf{x_i} - \mu_\mathbf{2})^\top \mathbf{\Sigma}^{-1})$

$= n/2\Sigma - 1/2 \Sigma_{i=1:y_i=1}^n (\mathbf{x_i} - \mu_\mathbf{1})(\mathbf{x_i} - \mu_\mathbf{1})^\top - 1/2 \Sigma_{i=1:y_i=2}^n (\mathbf{x_i} - \mu_\mathbf{2})(\mathbf{x_i} - \mu_\mathbf{2})^\top$

$\Rightarrow n\hat{\Sigma} - \hat{\Sigma}_{i=1:y_i=1}^n (\mathbf{x_i} - \hat{\mu}_\mathbf{1})(\mathbf{x_i} - \hat{\mu}_\mathbf{1})^\top - \Sigma_{i=1:y_i=2}^n (\mathbf{x_i} - \hat{\mu}_\mathbf{2})(\mathbf{x_i} - \hat{\mu}_\mathbf{2})^\top = 0$

$\Rightarrow \hat{\Sigma} = \frac{\Sigma_{i=1:y_i=1}^n (\mathbf{x_i} - \hat{\mu}_\mathbf{1})(\mathbf{x_i} - \hat{\mu}_\mathbf{1})^\top + \Sigma_{i=1:y_i=2}^n (\mathbf{x_i} - \hat{\mu}_\mathbf{2})(\mathbf{x_i} - \hat{\mu}_\mathbf{2})^\top}{n}$

$\Rightarrow \hat{\Sigma} = \frac{n_1 S_1 + n_2 S_2}{n}$

Question 2.4

$\log \frac{p(y_i=1|\mathbf{x_i})}{p(y_i=2|\mathbf{x_i})}$

$= \log \frac{p(y_i=1,\mathbf{x_i})}{p(y_i=2,\mathbf{x_i})}$

$= \log \frac{p(\mathbf{x_i}|y_i=1)p(y_i=1)}{p(\mathbf{x_i}|y_i=2)p(y_i=2)}$

$= \log p(\mathbf{x_i}|y_i = 1) - \log p(\mathbf{x_i}|y_i = 2) + \log \frac{p(y_i=1)}{p(y_i=2)}$

$= -d/2 \log(2\pi) - 1/2 \log |\hat{\mathbf{\Sigma}}| + \frac{-(\mathbf{x_i}-\hat{\mu}_\mathbf{1})^\top \hat{\mathbf{\Sigma}}^{-1}(\mathbf{x_i}-\hat{\mu}_\mathbf{1})}{2} + d/2 \log(2\pi) + 1/2 \log |\hat{\mathbf{\Sigma}}| - \frac{-(\mathbf{x_i}-\hat{\mu}_\mathbf{2})^\top \hat{\mathbf{\Sigma}}^{-1}(\mathbf{x_i}-\hat{\mu}_\mathbf{2})}{2} + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

$= -1/2(\mathbf{x_i}^\top - \hat{\mu}_\mathbf{1}^\top)\hat{\mathbf{\Sigma}}^{-1}(\mathbf{x_i} - \hat{\mu}_\mathbf{1}) + 1/2(\mathbf{x_i}^\top - \hat{\mu}_\mathbf{2}^\top)\hat{\mathbf{\Sigma}}^{-1}(\mathbf{x_i} - \hat{\mu}_\mathbf{2}) + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

$= -1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x_i} + 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{1} + 1/2\hat{\mu}_\mathbf{1}^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x_i} - 1/2\hat{\mu}_\mathbf{1}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{1} + 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x_i} - 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{2} - 1/2\hat{\mu}_\mathbf{2}^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x_i} + 1/2\hat{\mu}_\mathbf{2}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{2} + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

$= \hat{\mu}_\mathbf{1}^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x_i} - 1/2\hat{\mu}_\mathbf{1}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{1} + \hat{\mu}_\mathbf{2}^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x_i} + 1/2\hat{\mu}_\mathbf{2}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{2} + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

$= (\hat{\mu}_\mathbf{1}^\top + \hat{\mu}_\mathbf{2}^\top)\hat{\mathbf{\Sigma}}^{-1}\mathbf{x_i} - 1/2\hat{\mu}_\mathbf{1}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{1} + 1/2\hat{\mu}_\mathbf{2}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{2} + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

$= \mathbf{w}^\top \mathbf{x_i} + w_0 = 0$

Where:

$\mathbf{w} = (\hat{\mu}_\mathbf{1}^\top + \hat{\mu}_\mathbf{2}^\top)\hat{\mathbf{\Sigma}}^{-1}$

$w_0 = -1/2\hat{\mu}_\mathbf{1}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{1} + 1/2\hat{\mu}_\mathbf{2}^\top \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_\mathbf{2} + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

Therefore linear

Question 2.5

It follows from 2.2

$\Rightarrow \frac{n_1}{n_1+n_2} = \hat{\eta}$

$\Rightarrow \frac{\Sigma_{i=1:y_i=1}^n \mathbf{x_i}}{n_1} = \hat{\mu}_\mathbf{1}$

$\Rightarrow \frac{\Sigma_{i=1:y_i=2}^n \mathbf{x_i}}{n_2} = \hat{\mu}_\mathbf{2}$

$\frac{\partial}{\partial \mathbf{\Sigma_1}^{-1}} \Sigma_{i=1}^n \log p(\mathbf{x_i}, y_i)$

$= n_1/2 \frac{\partial}{\partial \mathbf{\Sigma_1}^{-1}} \log |\mathbf{\Sigma_1^{-1}}| - 1/2 \Sigma_{i=1:y_i=1}^n \frac{\partial}{\partial \mathbf{\Sigma_1}^{-1}} \frac{-(\mathbf{x_i}-\mu_\mathbf{1})^\top \mathbf{\Sigma_1}^{-1}(\mathbf{x_i}-\mu_\mathbf{1})}{2}$

$= n_1/2\Sigma_1 - 1/2\Sigma_{i=1:y_i=1}^n (\mathbf{x_i} - \mu_\mathbf{1})(\mathbf{x_i} - \mu_\mathbf{1})^\top$

$\Rightarrow \frac{\Sigma_{i=1:y_i=1}^n (\mathbf{x_i}-\mu_\mathbf{1})(\mathbf{x_i}-\mu_\mathbf{1})^\top}{n_1} = S_1 = \hat{\Sigma}_1$

Similar case follows as $\hat{\Sigma}_1$, $\frac{\Sigma_{i=1:y_i=2}^n (\mathbf{x_i}-\mu_\mathbf{2})(\mathbf{x_i}-\mu_\mathbf{2})^\top}{n_2} = S_2 = \hat{\Sigma}_2$

$\log \frac{p(y_i=1|\mathbf{x_i})}{p(y_i=2|\mathbf{x_i})}$

$= \log p(\mathbf{x_i}|y_i = 1) - \log p(\mathbf{x_i}|y_i = 2) + \log \frac{p(y_i=1)}{p(y_i=2)}$

$= -d/2 \log(2\pi) - 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{1}| + \frac{-(\mathbf{x_i}-\hat{\mu}_\mathbf{1})^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}(\mathbf{x_i}-\hat{\mu}_\mathbf{1})}{2} + d/2 \log(2\pi) + 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{2}| - \frac{-(\mathbf{x_i}-\hat{\mu}_\mathbf{2})^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}(\mathbf{x_i}-\hat{\mu}_\mathbf{2})}{2} + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

$= 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{2}| - 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{1}| - 1/2(\mathbf{x_i}^\top - \hat{\mu}_\mathbf{1}^\top)\hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}(\mathbf{x_i} - \hat{\mu}_\mathbf{1}) + 1/2(\mathbf{x_i}^\top - \hat{\mu}_\mathbf{2}^\top)\hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}(\mathbf{x_i} - \hat{\mu}_\mathbf{2}) + \log \frac{\hat{\eta}}{1-\hat{\eta}}$

$= 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{2}| - 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{1}| - 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mathbf{x_i} + 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mu_\mathbf{1} + 1/2\mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mathbf{x_i} - 1/2\mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mu_\mathbf{1} + 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mathbf{x_i} - 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mu_\mathbf{2} - 1/2\mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mathbf{x_i} + 1/2\mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mu_\mathbf{2}$

$= 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{2}| - 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{1}| - 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mathbf{x_i} + \mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mathbf{x_i} - 1/2\mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mu_\mathbf{1} + 1/2\mathbf{x_i}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mathbf{x_i} - \mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mathbf{x_i} + 1/2\mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mu_\mathbf{2}$

$= \mathbf{x_i}^\top(-1/2\hat{\mathbf{\Sigma}}_\mathbf{1}^{-1} + 1/2\hat{\mathbf{\Sigma}}_\mathbf{2}^{-1})\mathbf{x_i} + (\mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1} - \mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1})\mathbf{x_i} - 1/2\mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mu_\mathbf{1} + 1/2\mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mu_\mathbf{2} + 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{2}| - 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{1}|$

$= \mathbf{x_i}^\top \mathbf{W}\mathbf{x_i} + \mathbf{w}^\top \mathbf{x_i} + w_0 = 0$

Where:

$\mathbf{W} = -1/2\hat{\mathbf{\Sigma}}_\mathbf{1}^{-1} + 1/2\hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}$

$\mathbf{w}^\top = \mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1} - \mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}$

$w_0 = -1/2\mu_\mathbf{1}^\top \hat{\mathbf{\Sigma}}_\mathbf{1}^{-1}\mu_\mathbf{1} + 1/2\mu_\mathbf{2}^\top \hat{\mathbf{\Sigma}}_\mathbf{2}^{-1}\mu_\mathbf{2} + 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{2}| - 1/2 \log |\hat{\mathbf{\Sigma}}_\mathbf{1}|$

Therefore quadratic

Question 3

Question 3.1

```
top <- "/Users/vladislavtrukhin/Downloads/SpamAssassin"
Directories <- c("easy_ham","spam")
dirs <- paste(top, Directories, sep ="/")
source("/Users/vladislavtrukhin/Downloads/SpamAssassin/readRawEmail.R")
mail <- readAllMessages(dirs = dirs)
```

```
doc <- c()
for (i in 1:3184) {
  tmp <- mail[[i]]$body
  tmp2 <- paste(tmp$text,collapse="")
  r <- "\\b([[:punct:]|[:digit:]])*[a-zA-Z]*([[:punct:]|[:digit:]])+[a-zA-Z]*([[:punct:]|[:digit:]])*"
  tmp3 <- gsub(r," ",tmp2)
  tmp4 <- gsub("[^A-Za-z]"," ",tmp3)
  doc <- cbind(doc, tmp4)
}

corpus <- Corpus(VectorSource(doc))
res <- TermDocumentMatrix(corpus, control = list(removePunctuation = TRUE,
                                                  stemming = TRUE, wordLengths = c(3, 20)))
res <- as.matrix(res)
```

```
q1h <- rowSums(res[,1:2188]) / rowSums(res[,1:2188] > 0)
q2h <- rowSums(res[,1:2188] > 0) / ncol(res[,1:2188])

q1s <- rowSums(res[,2189:3184]) / rowSums(res[,2189:3184] > 0)
q2s <- rowSums(res[,2189:3184] > 0) / ncol(res[,2189:3184])

tail(sort(q1h),10) # Top 10 ham words with largest quantity 1
```

```
##       msgs standardis        the   dinosaur    dirksen      tribe      powel
##    14.0000    14.0000    14.3357    14.5000    15.0000    16.0000    17.0000
## friendship     maxlin     hextab
##    18.0000    19.0000    20.0000
```

```r
tail(sort(q2h),10) # Top 10 ham words with largest quantity 2
```

```
##       but       not       you      this      have      with       for      that
## 0.4867459 0.4908592 0.5063985 0.5420475 0.5470750 0.5489031 0.6512797 0.6681901
##       and       the
## 0.7838208 0.8999086
```

```r
tail(sort(q1s),10) # Top 10 spam words with largest quantity 1
```

```
##        les marshalles         des        wake     marshal        king        atol
##         27          28          33          33          34          44          59
##    enenkio      island     kingdom
##         79          82          90
```

```r
tail(sort(q2s),10) # Top 10 spam words with largest quantity 2
```

```
##      with       our       are      from      your       for      this       and
## 0.4497992 0.4779116 0.4909639 0.4909639 0.6084337 0.6094378 0.6345382 0.6375502
##       you       the
## 0.6606426 0.6817269
```

Question 3.2

```r
set.seed(1)

testingidx <- sample(1:ncol(res),100)
trainingidx <- 1:ncol(res)
trainingidx <- trainingidx[-testingidx]

# Sufficient statistics
y <- res
w <- res > 0

w_tr_hm <- w[,trainingidx[!trainingidx > 2188]]
w_tr_sp <- w[,trainingidx[trainingidx > 2188]]

y_tr_hm <- y[,trainingidx[!trainingidx > 2188]]
y_tr_sp <- y[,trainingidx[trainingidx > 2188]]

w_te <- w[,testingidx]
y_te <- y[,testingidx]

# Model fitting
lambda_hm <- rowSums(w_tr_hm*(y_tr_hm-1)) / rowSums(w_tr_hm)
lambda_hm[!is.finite(lambda_hm)] <- 0
theta_hm <- rowSums(w_tr_hm) / sum(!trainingidx > 2188)

lambda_sp <- rowSums(w_tr_sp*(y_tr_sp-1)) / rowSums(w_tr_sp)
lambda_sp[!is.finite(lambda_sp)] <- 0
theta_sp <- rowSums(w_tr_sp) / sum(trainingidx > 2188)

# Using model on testing data
log_hm <- log(sum(trainingidx > 2188)) - log(length(trainingidx))
log_sp <- log(sum(!trainingidx > 2188)) - log(length(trainingidx))
```

```r
log_ratio <- w_te*(log(theta_hm+0.0001) - log(theta_sp+0.0001) - lambda_hm + lambda_sp
                   + (y_te-1)*(log(lambda_hm+0.0001) - log(lambda_sp+0.0001)))
log_ratio <- log_ratio + (1-w_te)*(log(1-theta_hm) - log(1-theta_sp))
log_ratio <- colSums(log_ratio) + log_hm - log_sp

# Prediction accuracy on testing data
sum((log_ratio > 0) == (!testingidx > 2188)) / length(testingidx)
```

```
## [1] 0.97
```

Question 3.3

```r
doc <- c()
for (i in 1:3184) {
  tmp <- mail[[i]]$body
  tmp2 <- paste(tmp$text,collapse="")
  r <- "\\b([[:punct:]|[:digit:]])*[a-zA-Z]*([[:punct:]|[:digit:]])+[a-zA-Z]*([[:punct:]|[:digit:]])"
  tmp3 <- gsub(r," ",tmp2)
  tmp4 <- gsub("[^A-Za-z]"," ",tmp3)
  doc <- cbind(doc, tmp4)
}

corpus <- Corpus(VectorSource(doc))
res <- TermDocumentMatrix(corpus, control = list(removePunctuation = TRUE,
                                                  stemming = TRUE, wordLengths = c(3, 20)))
res <- as.matrix(res)
```

```r
set.seed(1)

testingidx <- sample(1:ncol(res),100)
trainingidx <- 1:ncol(res)
trainingidx <- trainingidx[-testingidx]

# Sufficient statistics
y <- res
w <- res > 0

w_tr_hm <- w[,trainingidx[!trainingidx > 2188]]
w_tr_sp <- w[,trainingidx[trainingidx > 2188]]

y_tr_hm <- y[,trainingidx[!trainingidx > 2188]]
y_tr_sp <- y[,trainingidx[trainingidx > 2188]]

w_te <- w[,testingidx]
y_te <- y[,testingidx]

# Model fitting
lambda_hm <- rowSums(w_tr_hm*(y_tr_hm-1)) / rowSums(w_tr_hm)
lambda_hm[!is.finite(lambda_hm)] <- 0
theta_hm <- rowSums(w_tr_hm) / sum(!trainingidx > 2188)

lambda_sp <- rowSums(w_tr_sp*(y_tr_sp-1)) / rowSums(w_tr_sp)
lambda_sp[!is.finite(lambda_sp)] <- 0
theta_sp <- rowSums(w_tr_sp) / sum(trainingidx > 2188)
```

```r
# Using model on testing data
log_hm <- log(sum(trainingidx > 2188)) - log(length(trainingidx))
log_sp <- log(sum(!trainingidx > 2188)) - log(length(trainingidx))

log_ratio <- w_te*(log(theta_hm+0.0001) - log(theta_sp+0.0001) - lambda_hm + lambda_sp
                   + (y_te-1)*(log(lambda_hm+0.0001) - log(lambda_sp+0.0001)))
log_ratio <- log_ratio + (1-w_te)*(log(1-theta_hm) - log(1-theta_sp))
log_ratio <- colSums(log_ratio) + log_hm - log_sp

# Prediction accuracy on testing data
sum((log_ratio > 0) == (!testingidx > 2188)) / length(testingidx)
```

```
## [1] 0.98
```

The prediction accuracy is higher using the new regex, which differs in that it preserves contractions unlike the old regex. Contractions hold predictive value which were filtered out under the old regex.