**Notes/Links/oldcode\_MLHW6**

```{r Support\_Vector\_Classifier}

#do this on the NH using the training dat from before and see if you get different answers

NH\_SVC = NH %>%

NH\_SVC$diabetes <- relevel(NH$diabetes, ref = "No") #Set No Diabetes as Reference Level

set.seed(100)

train\_indices\_SVC <- createDataPartition(y = NH\_SVC$diabetes,p = 0.7,list = FALSE)

training\_SVC <- NH\_SVC[train\_indices\_SVC,]

testing\_SVC <- NH\_SVC[-train\_indices\_SVC,]

modelLookup("svmLinear")

set.seed(100)

#Set 10-fold cross-validation. Note if you want predicted probabilities, you need to set class Probs=True

train\_control\_SVC <- trainControl(method = "cv", number = 10, classProbs = T)

#Train model. Note we are scaling data

svm\_diabetes <- train(diabetes ~ ., data = training\_SVC, method = "svmLinear", trControl = train\_control\_SVC, preProcess = c("center", "scale"))

svm\_diabetes #accuracy:0.8961806

#Incorporate different values for cost paramter(cp) bc this method won't tune the hyperparameters for us. The CP is how much misclassification the support vector will allow

svm\_caret\_d<-train(diabetes ~ ., data=training\_SVC, method="svmLinear", trControl=train\_control\_SVC, preProcess=c("center", "scale"), tuneGrid=expand.grid(C=seq(0.001,2, length=30)))

#Visualize accuracy versus values of C. This shows how the accuracy changes based on the level of the cost I chose.

plot(svm\_caret\_d)

#Obtain metrics of accuracy from training

confusionMatrix(svm\_caret\_d) #Accuracy (average) : 0.8962

```

```{r Support\_Vector\_Classifier try two}

modelLookup("svmLinear")

set.seed(100)

#Set 10-fold cross-validation. Note if you want predicted probabilities, you need to set class Probs=True

train\_controlSVC <- trainControl(method = "cv", number = 10, classProbs = T)

svm <- train(diabetes ~ ., data = train\_data, method = "svmLinear", trControl = train\_controlSVC, preProcess = c("center", "scale"))

svm

#Incorporate different values for cost paramter(cp) bc this method won't tune the hyperparameters for us. The CP is how much misclassification the support vector will allow

svm.caret <- train(diabetes ~ ., data = train\_data, method = "svmLinear", trControl = train\_controlSVC, preProcess = c("center", "scale"), tuneGrid = expand.grid(C = seq(0.001,2, length = 30)))

#Visualize accuracy versus values of C. This shows how the accuracy changes based on the level of the cost I chose.

plot(svm.caret)

#Obtain metrics of accuracy from training

confusionMatrix(svm.caret)

```

logistic\_control <- trainControl(method = "cv", number = 10, sampling = "down"), preProc = c("center","scale")

set.seed(100)

logistic <- train(diabetes ~ ., data = train\_data, method = "glm", family = "binomial", trControl = logistic\_control)

summary(logistic)

confusionMatrix(logistic) #Accuracy (average) : 0.7319