

# Review Stat. 452

Prof. Eric A. Suess

May 6, 2020

# Review

We have studied Statistical Learning/Machine Learning in this class.

We have learned about

- ▶ Supervised Learning
- ▶ Unsupervised Learning

# Supervised Learning Algorithms

kNN, Naive Bayes, Decision Trees, Boosted Decision Trees, Random Forests, SVMs, Neural Networks

Multiple Linear Regression, Logistic Regression

# Unsupervised Learning Algorithms

Market Basket Analysis using Rule, working with Transaction Data

Clustering using k-Means

## Hold-out Method

We have used the Hold-out Method to learn or models on data.

Remember that the data used in the Training and Test data set needs to be randomized.

See the code from Chapter 6.

```
> indx <- sample(1:nrow(insurance), as.integer(0.9*nrow(ins
```

```
> indx
```

```
> insurance_train <- insurance[indx,]
```

```
> insurance_test <- insurance[-indx,]
```

You should check that you have done this for your project!

# Review

For the final you should review the ideas presented in the class after the midterm. The primary things you should review are:

- ▶ Ch. 6 Trees, Random Forests
- ▶ Ch. 7 ANN and SVM
- ▶ Ch. 8 Market Basket Analysis
- ▶ Ch. 9 Clustering

# Final

The final on Monday will be open book, open notes, open code, open Google, closed neighbor.

You should ask me questions during the exam if you need help.

Your answers will be typed into an R notebook.