

Day 10 - K Nearest Neighbors and Evaluating Classification Models

Oct. 8, 2020



Administrative

- **Homework 3** will be assigned Friday 10/9 and due Friday 10/23
- **Midterm** will be given Thursday 10/29 in class

From Pre-Class Assignment

Useful Stuff

- Videos were useful, but they were a little long
- I have a better idea of how we are evaluating classification models

Challenging bits

- There's so much terminology, do I have to remember it all?
- I'm still confused about the ROC and what it is doing.
- How is KNN a binary classifier?

The Confusion Matrix

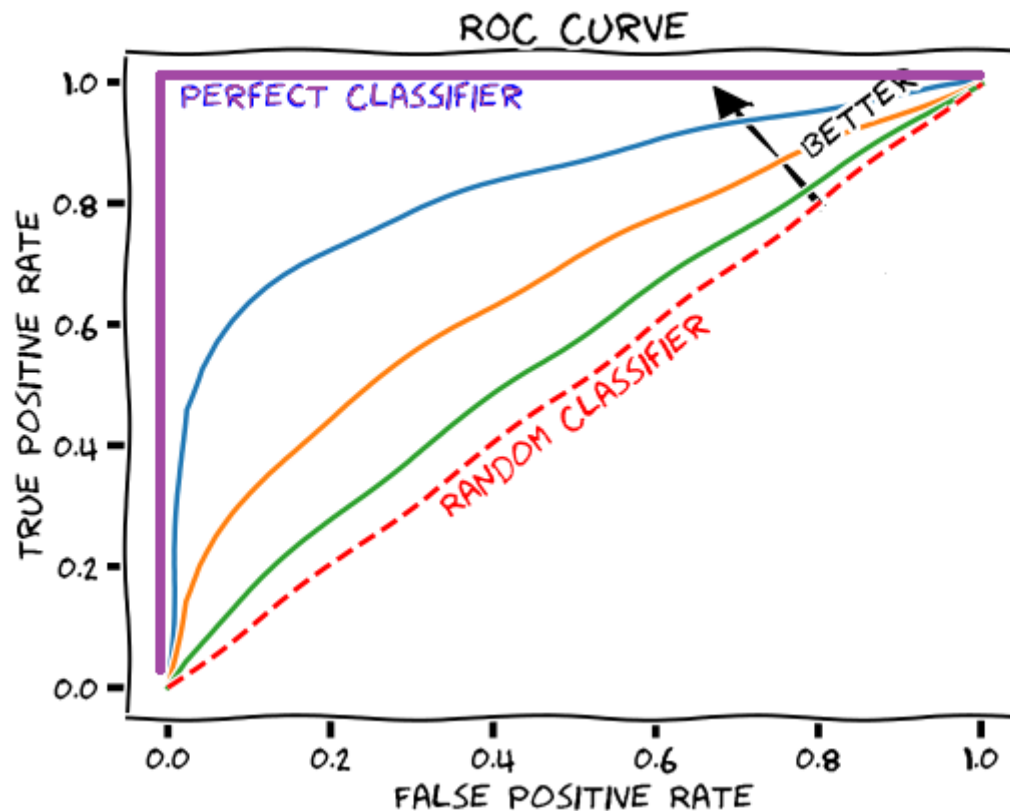
		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

```
from sklearn.metrics import confusion_matrix  
confusion_matrix(y_true, y_predicted)
```

Other Metrics

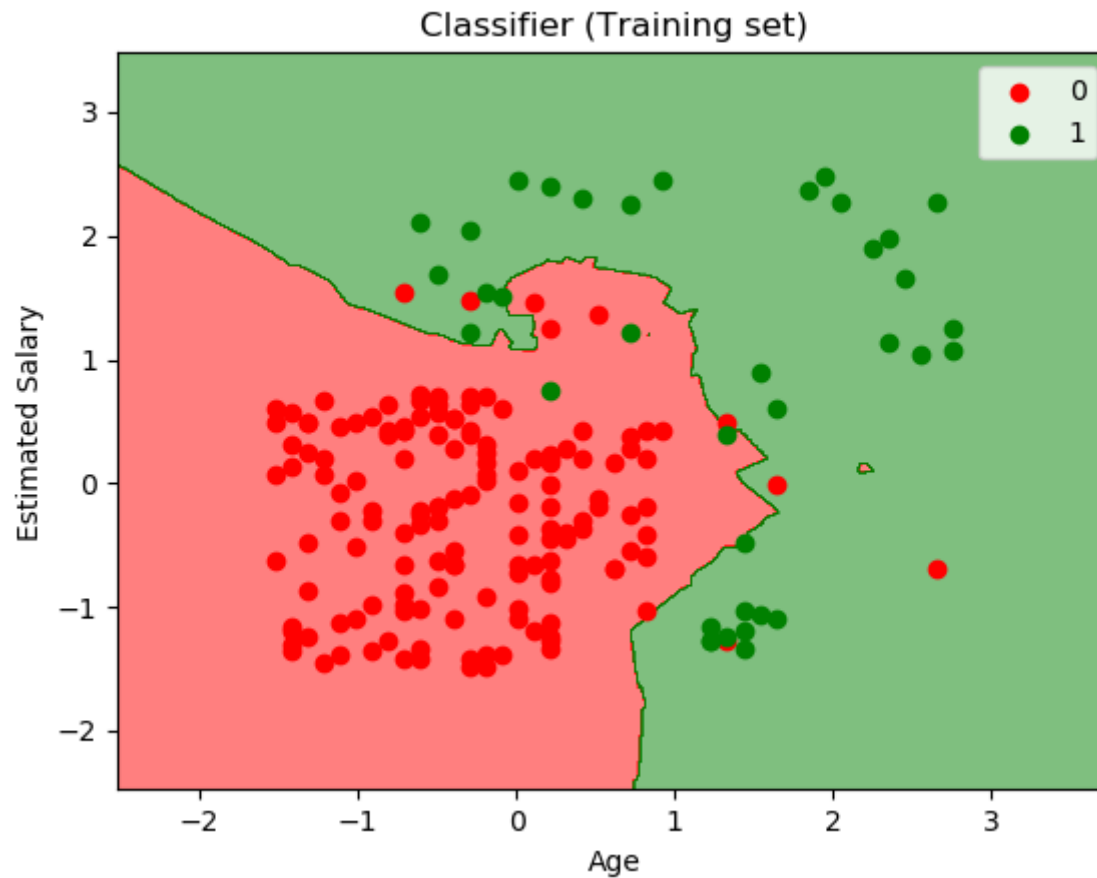
- Sensitivity (Recall): The ratio of True Positives to all True Cases $\frac{TP}{TP + FN}$
- Specificity: The ratio of True Negatives to all True Cases $\frac{TN}{TN + FP}$
- Precision: The ratio of True Positives to all Predicted Positives: $\frac{TP}{TP + FP}$
- F_1 Score: A balanced measure (0 to 1) that includes sensitivity and recall:
$$\frac{2TP}{2TP + FP + FN}$$

ROC Curve and AUC



```
from sklearn import metrics
fpr, tpr, thresholds = metrics.roc_curve(y_true, y_predict)
roc_auc = metrics.auc(fpr, tpr)
plt.plot(fpr, tpr)
```

KNN as a Binary Classifier



A Heads Up for Today

Working with Pima Diabetes Database, which has problems (zeros for various entries). It will read in fine, but you will need to replace zeros with something (not drop them).

For Section 2.2 (Imputing Data)

- Split data frame into positive and negative cases using the `label` as a mask
 - e.g., `pos = df[df.label == 1]`
- Replace zeros in a given columns with the `nan` (keeps zero from coming into the mean)
 - e.g., `pos['glucose'].replace(0, np.nan, inplace=True)`
- Replace `nan` in a given column with the mean
 - e.g., `pos['glucose'].fillna(pos['glucose'].mean())`
- Recombine positive and negative cases into a single data frame using `pd.concat()`

Questions, Comments, Concerns?