Classification and Regression Evaluation

	actual +	actual −
predicted +	TP	FP
predicted −	FN	TN

(confusion matrix)

accuracy = $\frac{TP + TN}{TP + TN + FP + FN}$

misclass. rate = 1 − accuracy

precision = $\frac{TP}{TP + FP}$ (% of correct + preds)

recall / TP rate = $\frac{TP}{TP + FN}$ (accuracy of + class)

~~TPR~~ (−)

TN rate = $\frac{TN}{TN + FP}$ (accuracy of − class)

note that for both regression + classification, you can always compare the likelihood on a test set (if it is a probabilistic model) to select amongst models. The abs. value of this is not meaningful, however.

0^{th} order comparison:

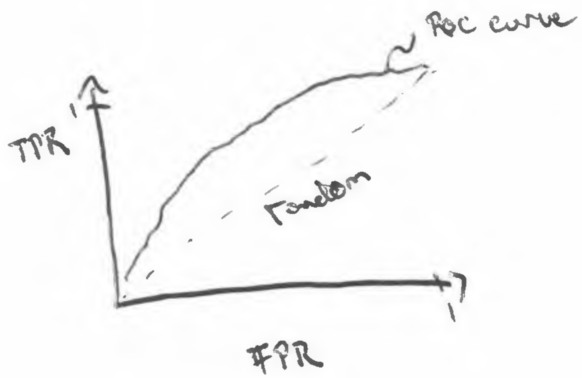
Compare model to a model that simply predicts most frequent class (w/o looking at input)

F_1 score = compromise between precision and recall

FN rate = $\frac{FN}{TP + FN}$ (error of + class)

FP rate = $\frac{FP}{TN + FP}$ (error of − class)

ROC curves show full spectrum of different TPR / FPR for different threshold values



- a single metric from this curve is the AUC (area under the curve)
  - 1 is the best.
  - .5 is random

Regression

0th order is to compare your model to just predicting the mean (this is basically r^2)

$$r^2 = 1 - \frac{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - \bar{y})^2}$$

$= 1 - \text{FUV}$

- (the mean squared error in units of the most simple model you can think of, just predicting a mean. 1 − to get $r^2 \in [\underset{\text{worse}}{0}, \underset{\text{better}}{1}]$) neg r^2 is very bad (worse than predicting just mean)

- also sometimes described as the fraction of explained variance

- MSE and RMSE

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$$

$$RMSE = \sqrt{MSE}$$

← nice because it expresses the error in the original units of y, so if "typical" values of y are y_{typ}, then we would like $RMSE < y_{typ}$.

- Several metrics are used to quantify performance ~~on the~~ using just the training set data. These metrics need to penalize somehow for # of features used in model bc more features typically means more overfitting.
  - adjusted r^2 : hacky way of penalizing the r^2 value for # of features
  - AIC : information theoretic criterion
  - BIC : derived from a bayesian POV