



Measuring Performance

CSCI 447/547 MACHINE LEARNING



Outline

- Confusion Matrix
 - F₁ Score
 - Gain and Lift Charts
 - Kolmogorov Smirnov Chart
 - ROC / AUC
 - Regression Metrics
 - Kappa Statistic
- 

Confusion Matrix

Confusion Matrix		Actual			
		Positive	Negative		
Predict	Positive	a	b	Precision	$a/(a+b)$
	Negative	c	d	Negative Predictive Value	$d/(d+c)$
		Sensitivity / Recall	Specificity	Accuracy = $(a+d)/(a+b+c+d)$	
		$a/(a+c)$	$d/(d+b)$		

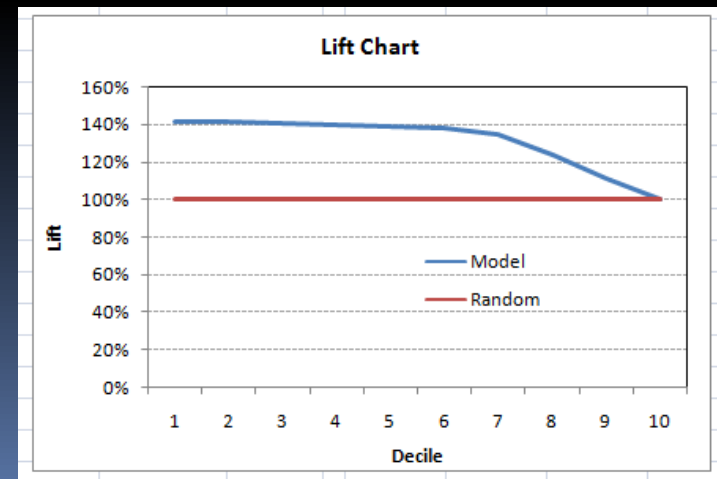
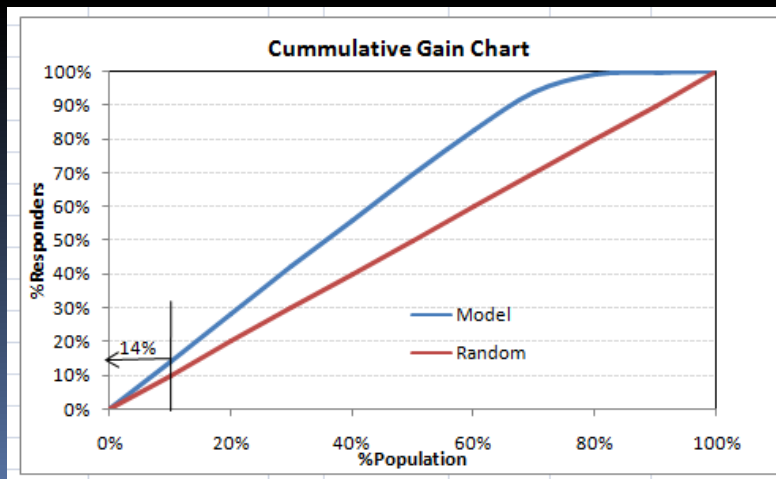
Confusion Matrix		Actual			
		1	0		
Predict	1	3,384	639	Precision	85.7%
	0	16	951	Negative Predictive Value	98.3%
		Sensitivity / Recall	Specificity	Accuracy = 88%	
		99.6%	59.8%		

F1 Score

- Good F1 score means you have low false positives and low false negatives
- $F_1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$
- Ranges from 0 to 1
- Higher values are better

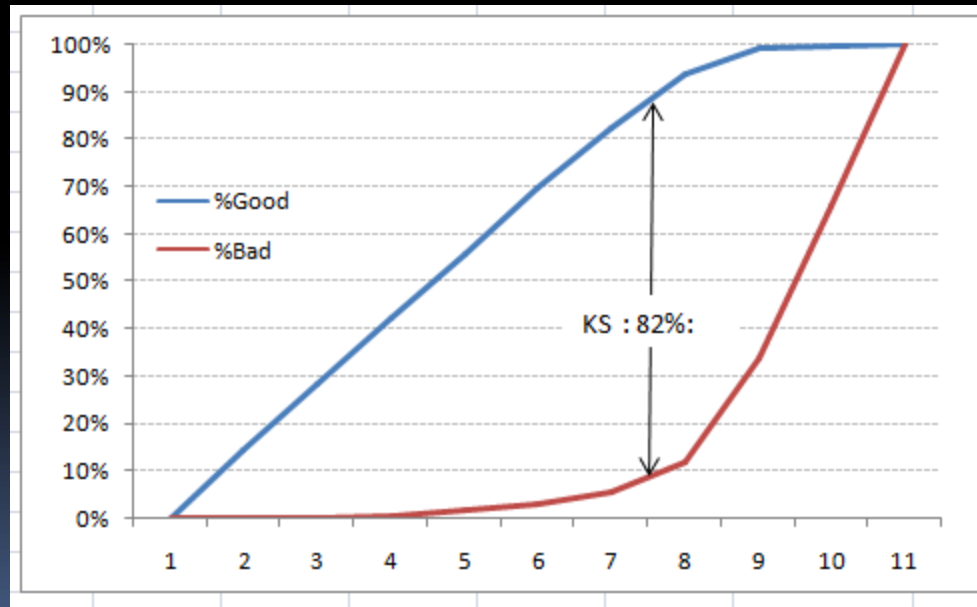
Gain and Lift Charts

- Calculate probability for each observation
- Sort in descending order
- Split into 10 partitions (deciles)
- Calculate correct predictions for each partition



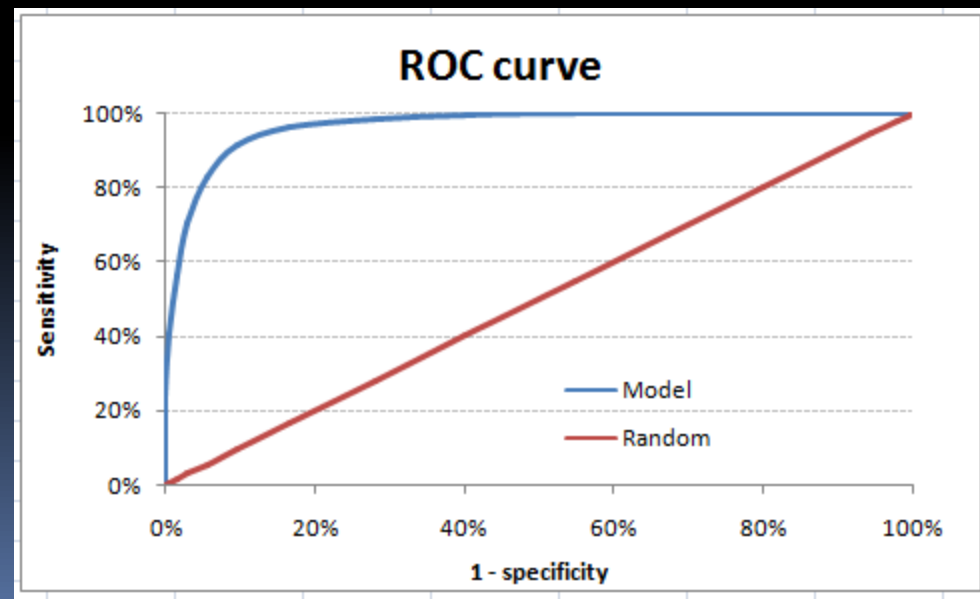
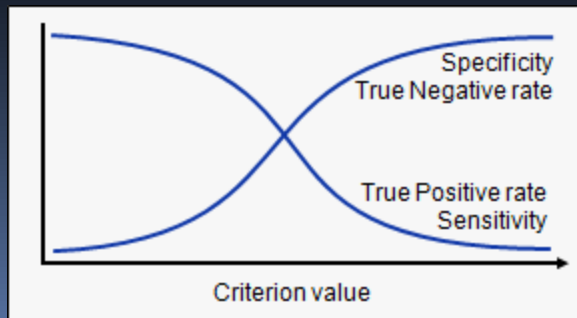
Kolmogorov Smirnov Chart

- Measure of the degree of separation between positive and negative distributions



ROC / AUC Curves

- Advantage over lift charts is that ROC is (almost) independent of the (possibly fluctuating) accuracy rate
- Measures model's ability to discriminate between positive and negative classes



Regression Metrics

- Mean Absolute Error
 - Gives an idea of magnitude of error but not direction
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
 - Converts MSE back to original magnitude
- R^2 (and Adjusted R^2)
 - Indication of correlation of predictions to actual values
 - Range between 0 and 1 with higher being better

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{Predicted}_i - \text{Actual}_i)^2}{N}}$$

Logarithmic Loss (Logloss)

- Evaluating the predictions of probabilities of membership in a given class
- Smaller is better, 0 is perfect

Kappa Statistic

- Cohen's Kappa
 - How much better than chance a model is
 - Range is -1 to 1, with higher being better
 - Some advise against using this
 - Dependent on distribution of correct and incorrect predictions, Cohen's Kappa can be misleading
- Power's Kappa (Informedness)
 - Likelihood of making an informed decision over a random guess
 - $\text{Recall} + \text{TNR} - 1$
 - Range is -1 to 1
 - 1: model is always correct; 0: model is random; -1: model is always incorrect

Summary

- Confusion Matrix
- F₁ Score
- Gain and Lift Charts
- Kolmogorov Smirnov Chart
- ROC / AUC
- Regression Metrics
- Kappa Statistic

