METIS

CLASSIFICATION ERROR METRICS

Supervised Learning

Regression

Classification

Supervised Learning

Regression

Classification

Supervised Learning

Classification

Predicting a class (one category, or many)

Examples

- Churn prediction
- Fraud detection

Supervised Learning

Classification: model examples

Logistic Regression

SVMs

Random Forest

KNNS

Identifying Problems

Which of your data science projects are supervised problems?

Of those, which are classification problems? What are the classes?

How do you determine which models work best?

Learning Objectives & Agenda

METIS

Learning objectives

Be able to

- Choose an appropriate error metric
- Evaluate a classifier’s performance
- Generalize error metrics to multiclass classification

Agenda

- Motivating Example
- Metrics
- Plots
- Multiclass problems

Motivating Example

METIS

Choosing the Right Error Measurement

You are asked to build a classifier for leukemia.

Your data set is *imbalanced:* 1% patients with leukemia, 99% healthy

You're asked to create a classifier with high *accuracy:* total % of predictions that are correct.

Choosing the Right Error Measurement

You are asked to build a classifier for leukemia.

Your data set is *imbalanced:* 1% patients with leukemia, 99% healthy

You're asked to create a classifier with high *accuracy:* total % of predictions that are correct.

Solution:

Build a simple model that always predicts "healthy". Accuracy will be 99%...

METRICS

METIS

Classification Metrics: Overview

Classification Performance

Confusion Matrix

Accuracy

Precision

Recall

F Score

ROC-AUC score

Performance Metrics

Classification: Confusion Matrix

	Actual Yes	Actual No
Predicted Yes	True Positive	False Positive
Predicted No	False Negative	True Negative

Performance Metrics

Classification: Confusion Matrix

	Actual Yes	Actual No
Predicted Yes	True Positive	False Positive
Predicted No	False Negative	True Negative

$$\text{True Positive Rate} = \frac{True\ Positive}{Actual\ Yes} \qquad \text{False Positive Rate} = \frac{False\ Positive}{Actual\ No}$$

$$\text{False Negative Rate} = \frac{False\ Negative}{Actual\ Yes} \qquad \text{True Negative Rate} = \frac{True\ Negative}{Actual\ No}$$

Performance Metrics

Classification: Confusion Matrix

	Actual Yes	Actual No
Predicted Yes	25	10
Predicted No	15	50

True Positive Rate $= \frac{25}{40}$

False Positive Rate $= \frac{10}{60}$

False Negative Rate $= \frac{15}{40}$

True Negative Rate $= \frac{50}{60}$

Performance Metrics

Classification: Accuracy

	Actual Yes	Actual No
Predicted Yes	True Positive	False Positive
Predicted No	False Negative	True Negative

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{All}}$$

Performance Metrics

Classification: Recall

	Actual Yes	Actual No
Predicted Yes	True Positive	False Positive
Predicted No	False Negative	True Negative

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

How well do you "recall" the targets?

Performance Metrics

Classification: Precision

	Actual Yes	Actual No
Predicted Yes	True Positive	False Positive
Predicted No	False Negative	True Negative

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

How "precise" are your predictions?

Performance Metrics

Classification: F Score

	Actual Yes	Actual No
Predicted Yes	True Positive	False Positive
Predicted No	False Negative	True Negative

$$\text{F Score} = 2\,\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Performance Metrics

Classification:

Metric	Meaning	Meaning
Accuracy	Percent accurately guessed	Higher is Better
Precision	Percent of guessed positive are positive	Higher is Better
Recall	Percent of positive were guessed positive	Higher is Better
F Score	Balances precision and recall	Higher is Better

Performance Metrics

Classification:

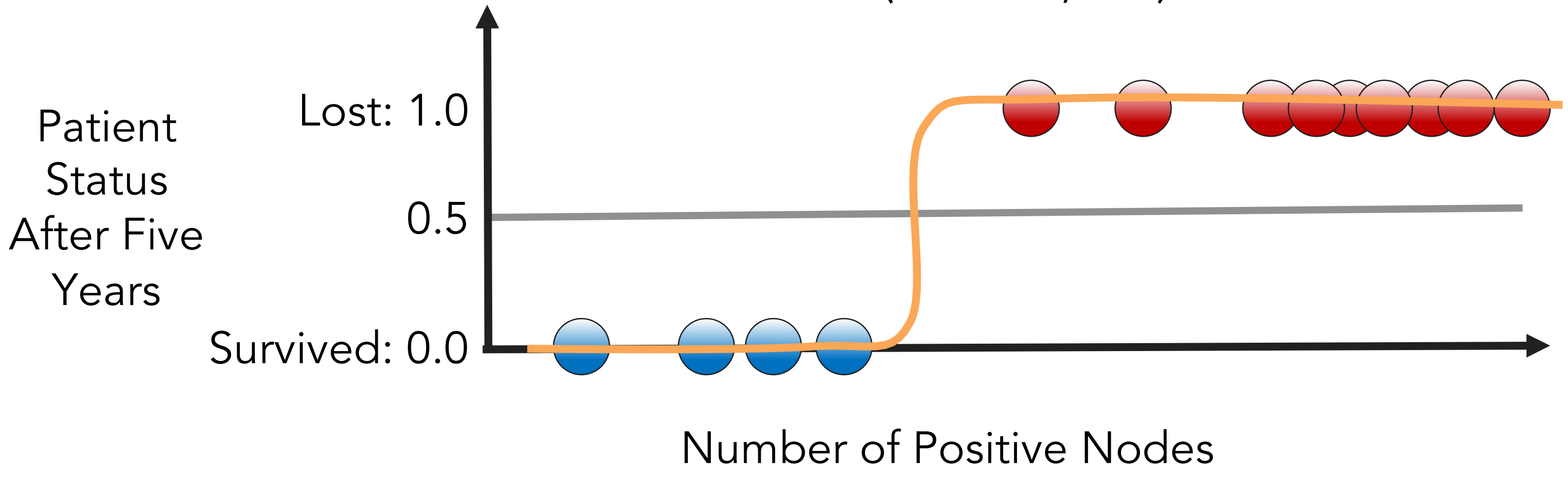
Metric	Why would you choose?
Accuracy	False positives and false negative are comparably costly
Precision	False positives are more costly that false negatives
Recall	False negatives are more costly than false positives
F Score	False positives and false negative are comparably costly. And data is unbalanced (low proportion of positive cases).

PLOTS

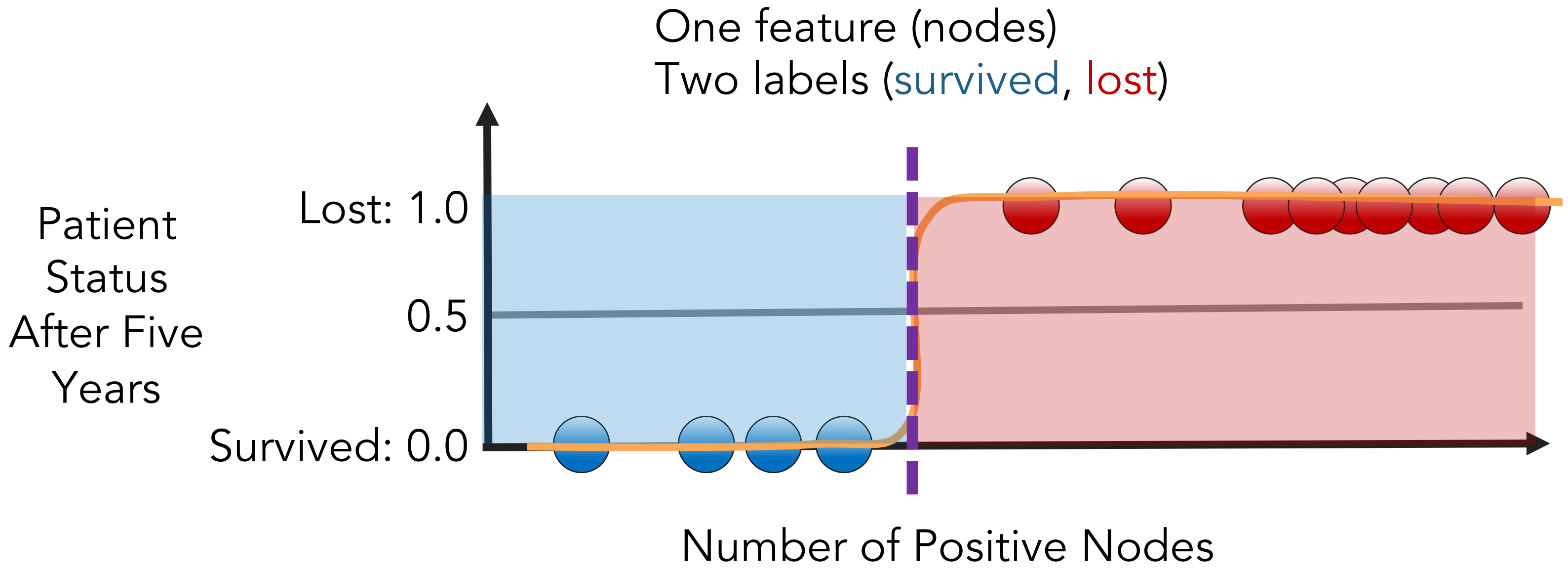
METIS

Classification Threshold

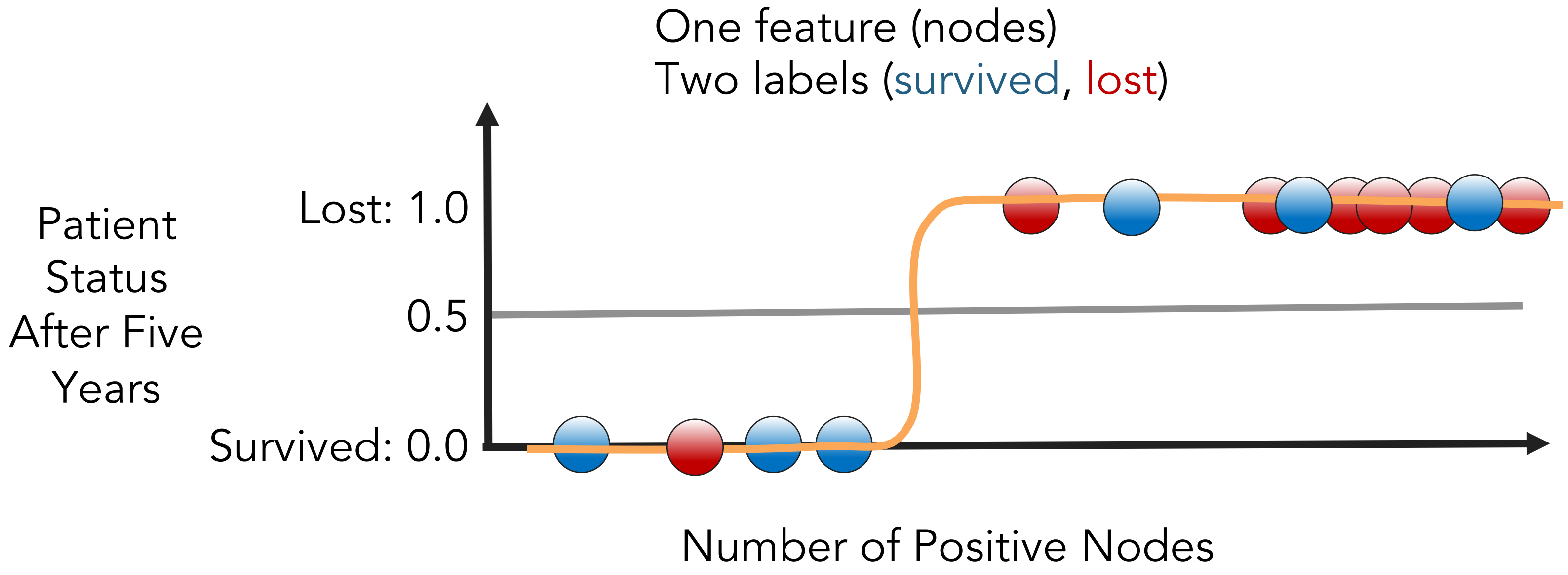
One feature (nodes)
Two labels (survived, lost)



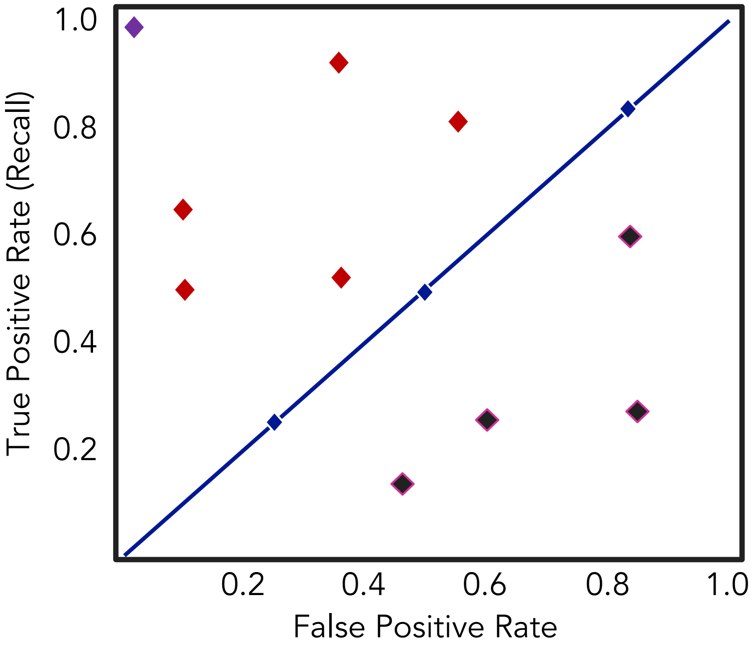
The Decision Boundary



The Decision Boundary

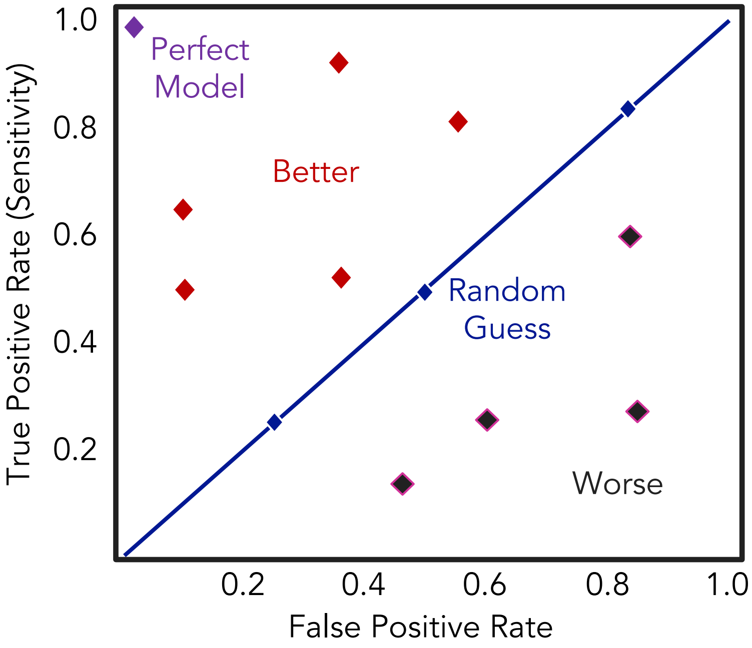


Receiver Operating Characteristic (ROC)



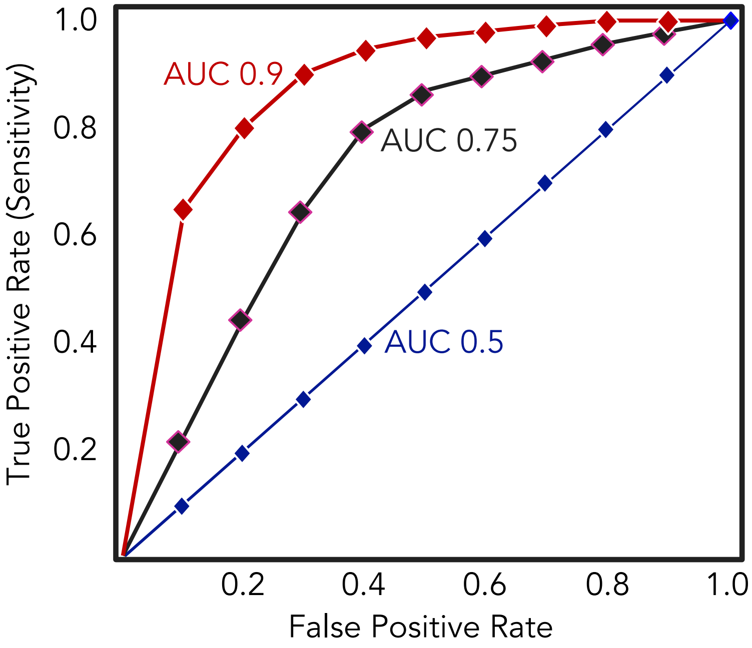
Evaluation of model at all possible thresholds

Receiver Operating Characteristic (ROC)



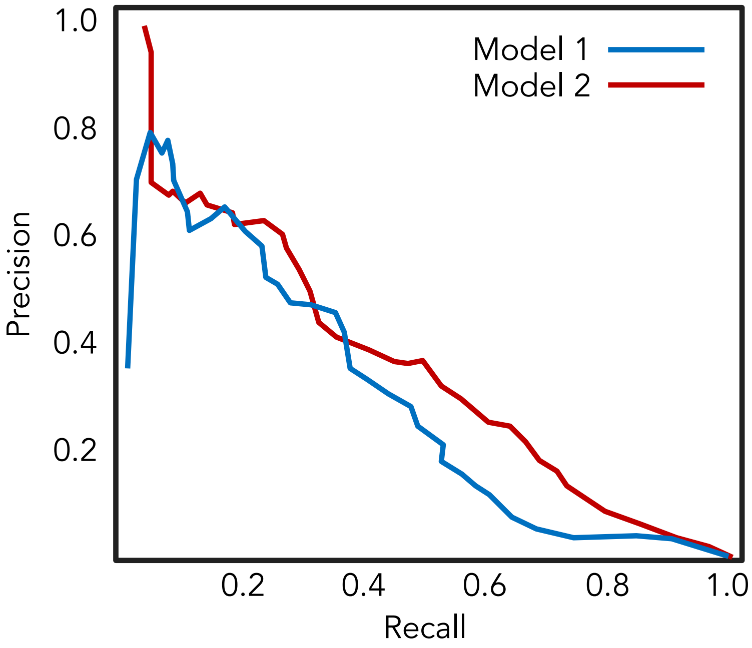
Evaluation of model at all possible thresholds

Area Under Curve (AUC)



Measures total area under ROC curve

Precision Recall Curve (PR Curve)



Measures trade-off between precision and recall

MULTI-CLASS

METIS

Multiple Class Error Metrics

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	TP1		
Actual Class 2		TP2	
Actual Class 3			TP3

Multiple Class Error Metrics

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	TP1		
Actual Class 2		TP2	
Actual Class 3			TP3

$$\text{Accuracy} = \frac{TP1 + TP2 + TP3}{\text{Total}}$$

Multiple Class Error Metrics

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	TP1		
Actual Class 2		TP2	
Actual Class 3			TP3

$$\text{Accuracy} = \frac{TP1 + TP2 + TP3}{\text{Total}}$$

Most multi-class error metrics are similar to binary versions—just expand elements as a sum

Recap

METIS

Learning objectives

We have discussed:

- Choose an appropriate error metric
- Evaluate a classifier's performance
- Generalize error metrics to multiclass classification

Takeaways

Choose an appropriate error metric for the business problem (*not* for the results)

Accuracy, precision, recall, specificity, and F1 are tailored for different needs

AUC measures how well two classes are being separated

Error metrics generalize to multiclass problems

QUESTIONS?

METIS

Performance Metrics

Regression

Adjusted R^2

MSE

MAE

F-Statistic

Likelihood