
FOR INSTRUCTOR PURPOSES ONLY

INSTRUCTOR NOTES

► Insert Text Here

FOR INSTRUCTOR PURPOSES ONLY

MATERIALS

► Insert Text Here

FOR INSTRUCTOR PURPOSES ONLY

PRE-WORK

► Insert Text Here

COMMUNICATING RESULTS

Chris Connell

COMMUNICATING RESULTS

LEARNING OBJECTIVES

- ▶ Explain the trade-offs between the precision and recall of a model while articulating the cost of false positives vs. false negatives
- ▶ Describe the difference between visualization for presentations vs. exploratory data analysis
- ▶ Know ROC/AUC and other measures or defining accuracy and when to use each

COURSE

PRE-WORK

PRE-WORK REVIEW

- ▶ Understand results from a confusion matrix and measure true positive rate and false positive rate
- ▶ Create and interpret results from a binary classification problem
- ▶ Know what a ROC and AUC are

INTRODUCTION

ADVANCED CLASSIFICATION METRICS

ADVANCED CLASSIFICATION METRICS

- ▶ Accuracy is only one of several metrics used when solving a classification problem.
- ▶ Accuracy = total predicted correct / total observations in dataset
- ▶ Accuracy alone doesn't always give us a full picture.
- ▶ If we know a model is 75% accurate, it doesn't provide *any* insight into why the 25% was wrong.

ADVANCED CLASSIFICATION METRICS

- ▶ Was it wrong across all labels?
- ▶ It's important to look at other metrics to fully understand the problem.

ADVANCED CLASSIFICATION METRICS

- ▶ We can split up the accuracy of each label by using the *true positive rate* and the *false positive rate*.
- ▶ For each label, we can put it into the category of a true positive, false positive, true negative, or false negative.

		<u>True class</u>	
		p	n
<u>Hypothesized class</u>	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

ADVANCED CLASSIFICATION METRICS

- ▶ True Positive Rate (TPR) asks, “Out of all of the target class labels, how many were accurately predicted to belong to that class?”
- ▶ For example, given a medical exam that tests for cancer, how often does it correctly identify patients with cancer?

		<u>True class</u>	
		p	n
<u>Hypothesized class</u>	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

$$TPR = TP/P = TP/(TP + FN)$$

ADVANCED CLASSIFICATION METRICS

- ▶ False Positive Rate (FPR) asks, “Out of all items not belonging to a class label, how many were predicted as belonging to that target class label?”
- ▶ For example, given a medical exam that tests for cancer, how often does it trigger a “false alarm” by incorrectly saying a patient has cancer?

		<u>True class</u>	
		p	n
<u>Hypothesized class</u>	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

$$\text{fp rate} = \frac{FP}{FP + TN}$$

ADVANCED CLASSIFICATION METRICS

- ▶ These can also be inverted.
- ▶ How often does a test *correctly* identify patients without cancer?

		<u>True class</u>	
		p	n
<u>Hypothesized class</u>	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

$\frac{TN}{TN+FN}$

ADVANCED CLASSIFICATION METRICS

- How often does a test *incorrectly* identify patient as cancer-free?

		<u>True class</u>	
		p	n
<u>Hypothesized class</u>	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

$\frac{FN}{TN+FN}$

ADVANCED CLASSIFICATION METRICS

- ▶ The true positive and false positive rates gives us a much clearer pictures of where predictions begin to fall apart.
- ▶ This allows us to adjust our models accordingly.

ADVANCED CLASSIFICATION METRICS

- ▶ A good classifier would have a true positive rate approaching 1 and a false positive rate approaching 0.
- ▶ In our smoking problem, this model would accurately predict *all* of the smokers as smokers and not accidentally predict any of the nonsmokers as smokers.

ADVANCED CLASSIFICATION METRICS

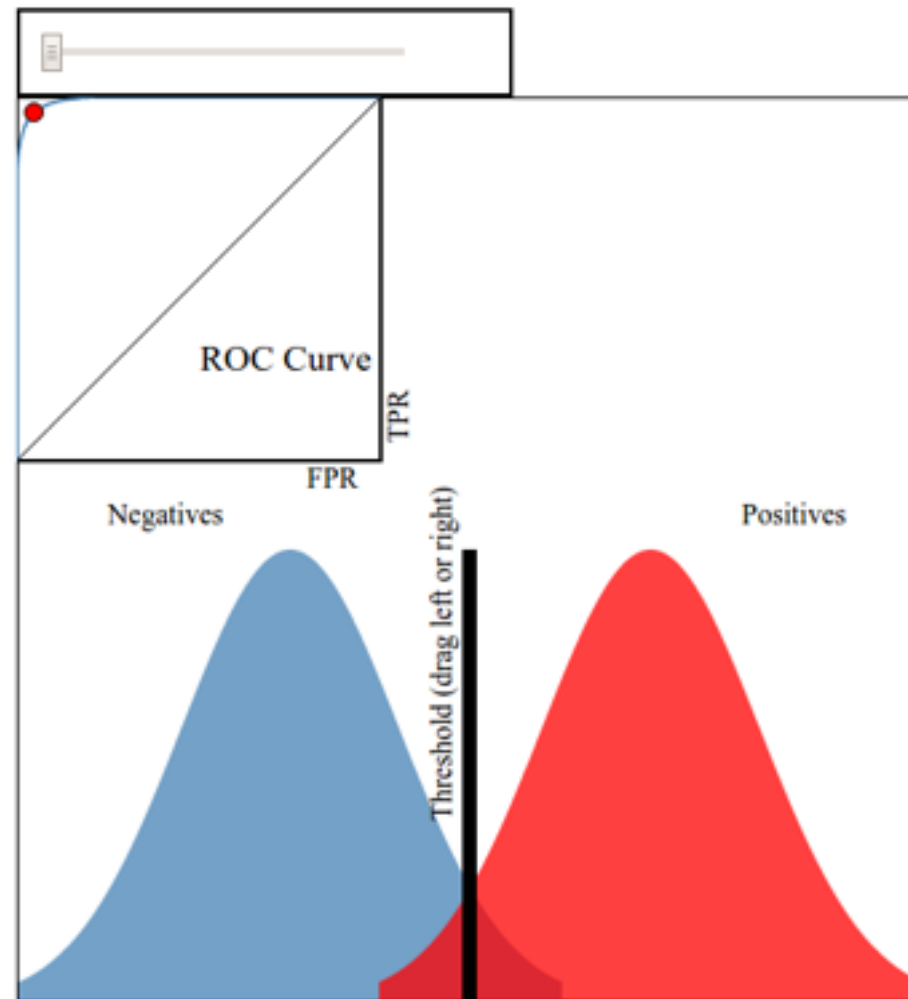
- ▶ We can vary the classification threshold for our model to get different predictions. But how do we know if a model is better overall than other model?
- ▶ We can compare the FPR and TPR of the models, but it can often be difficult to optimize two numbers at once.
- ▶ Logically, we like a single number for optimization.

ADVANCED CLASSIFICATION METRICS

- ▶ This is where the Receiver Operation Characteristic (ROC) curve comes in handy.
- ▶ The curve is created by plotting the true positive rate against the false positive rate at various model threshold settings.
- ▶ Area Under the Curve (AUC) summarizes the impact of TPR and FPR in one single value.

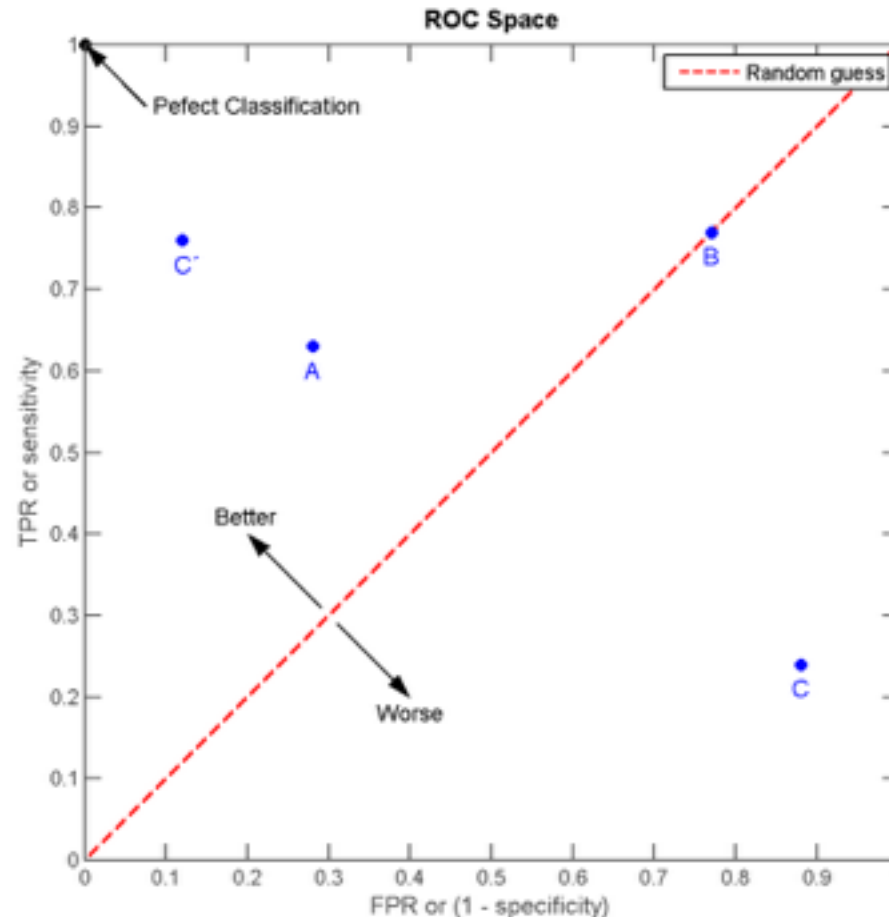
ADVANCED CLASSIFICATION METRICS

- ▶ This [interactive visualization](#) can help practice visualizing ROC curves.



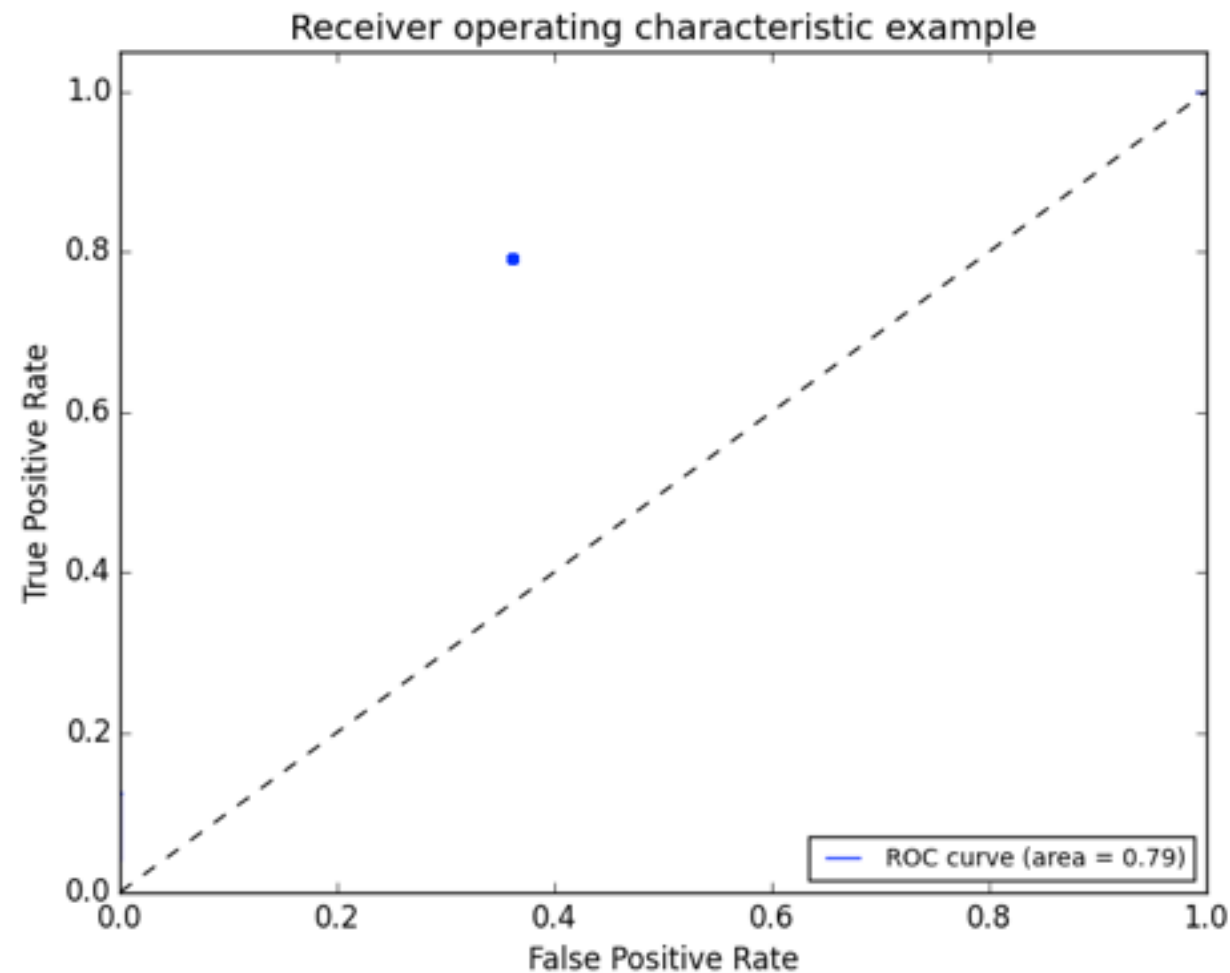
ADVANCED CLASSIFICATION METRICS

- There can be a variety of points on an ROC curve.



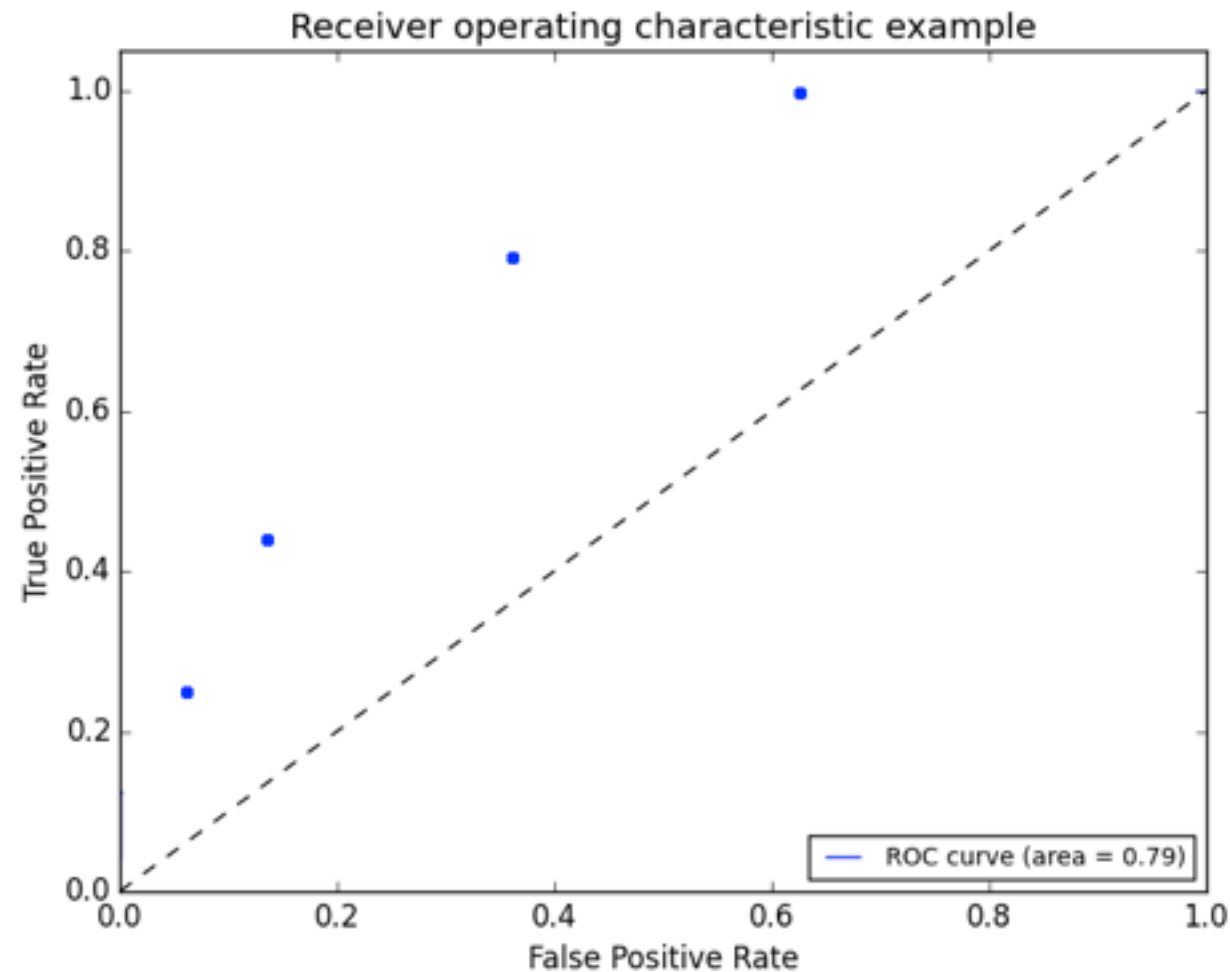
ADVANCED CLASSIFICATION METRICS

- ▶ We can begin by plotting an individual TPR/FPR pair for one threshold.



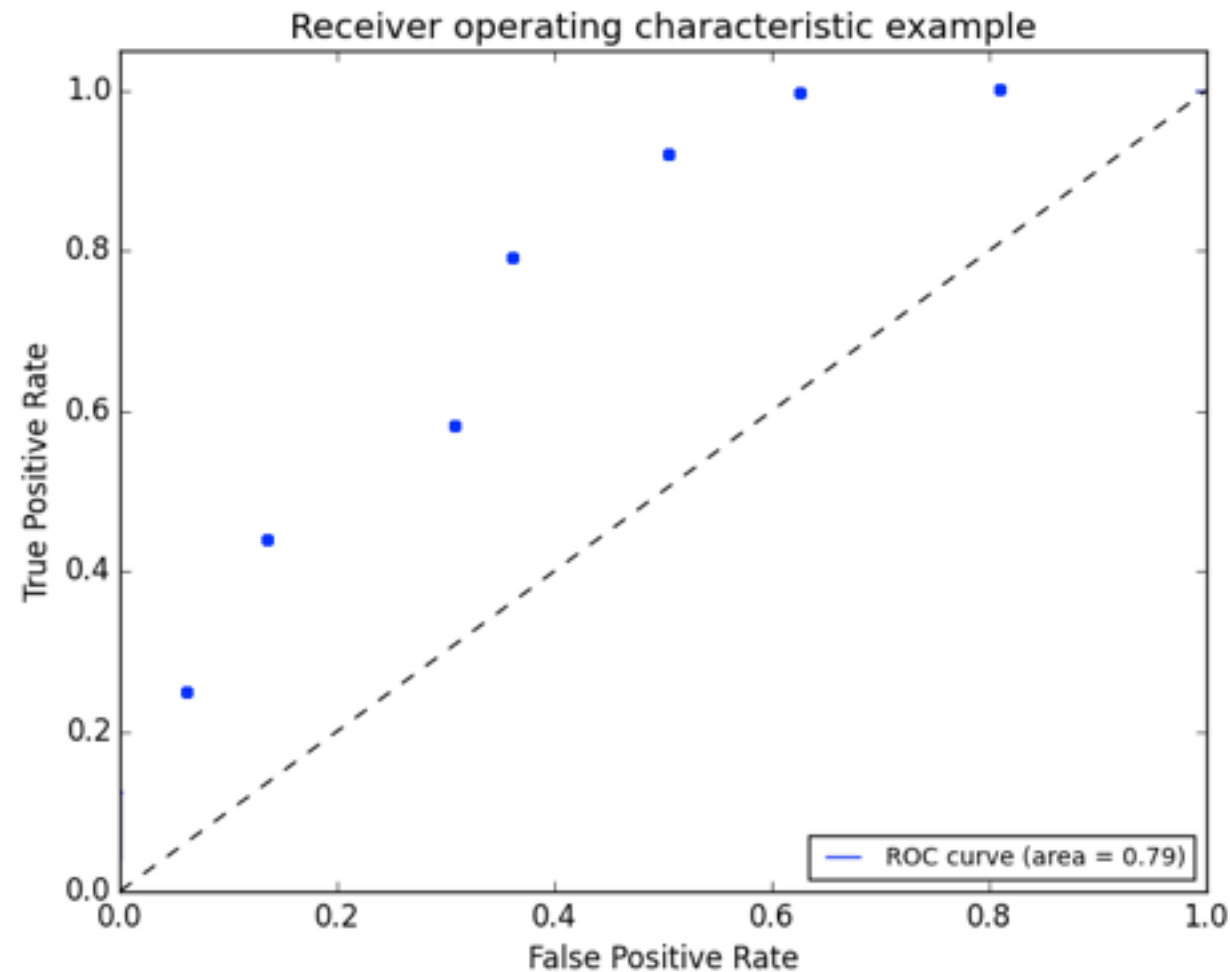
ADVANCED CLASSIFICATION METRICS

- We can continue adding pairs for different thresholds



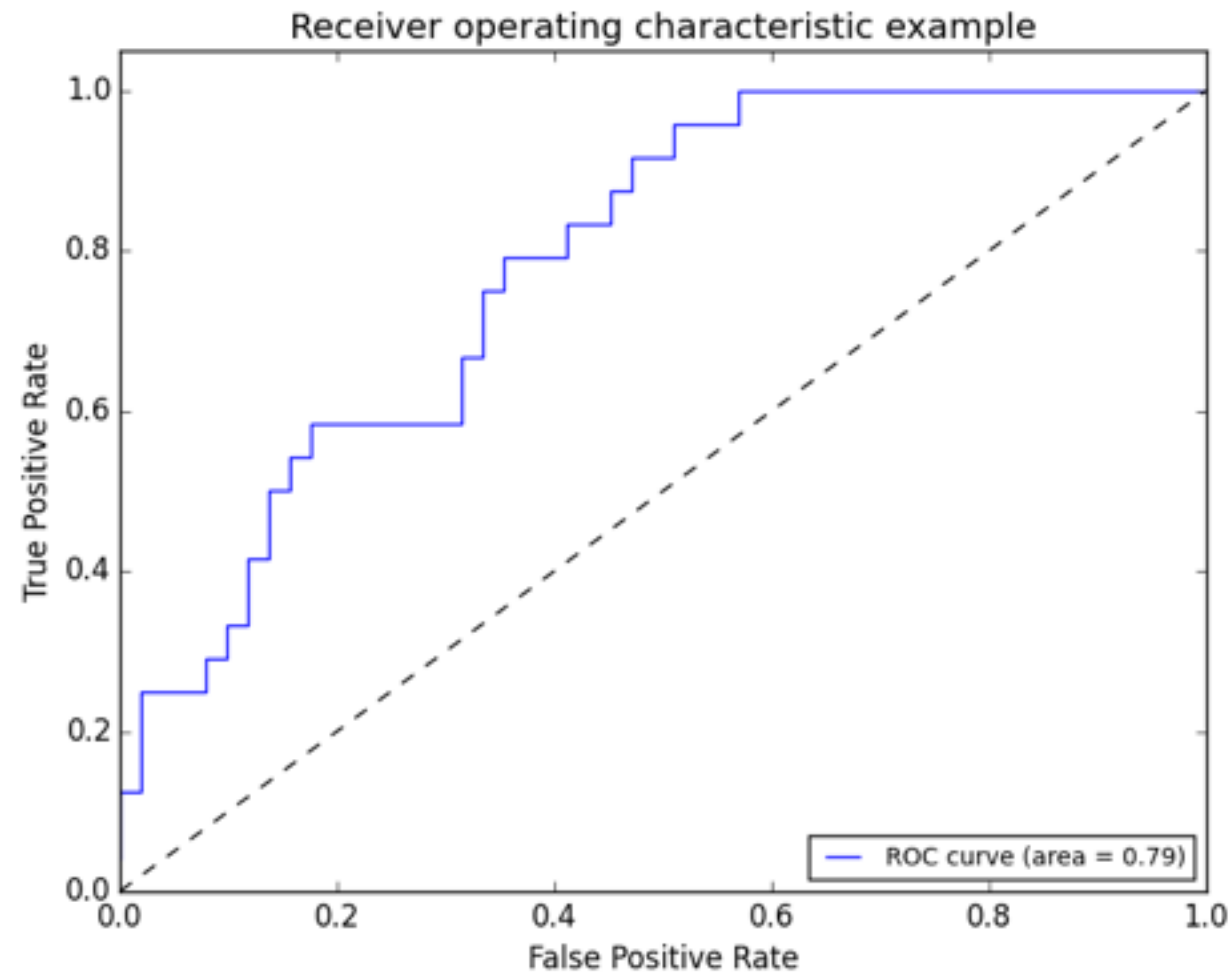
ADVANCED CLASSIFICATION METRICS

- ▶ We can continue adding pairs for different thresholds



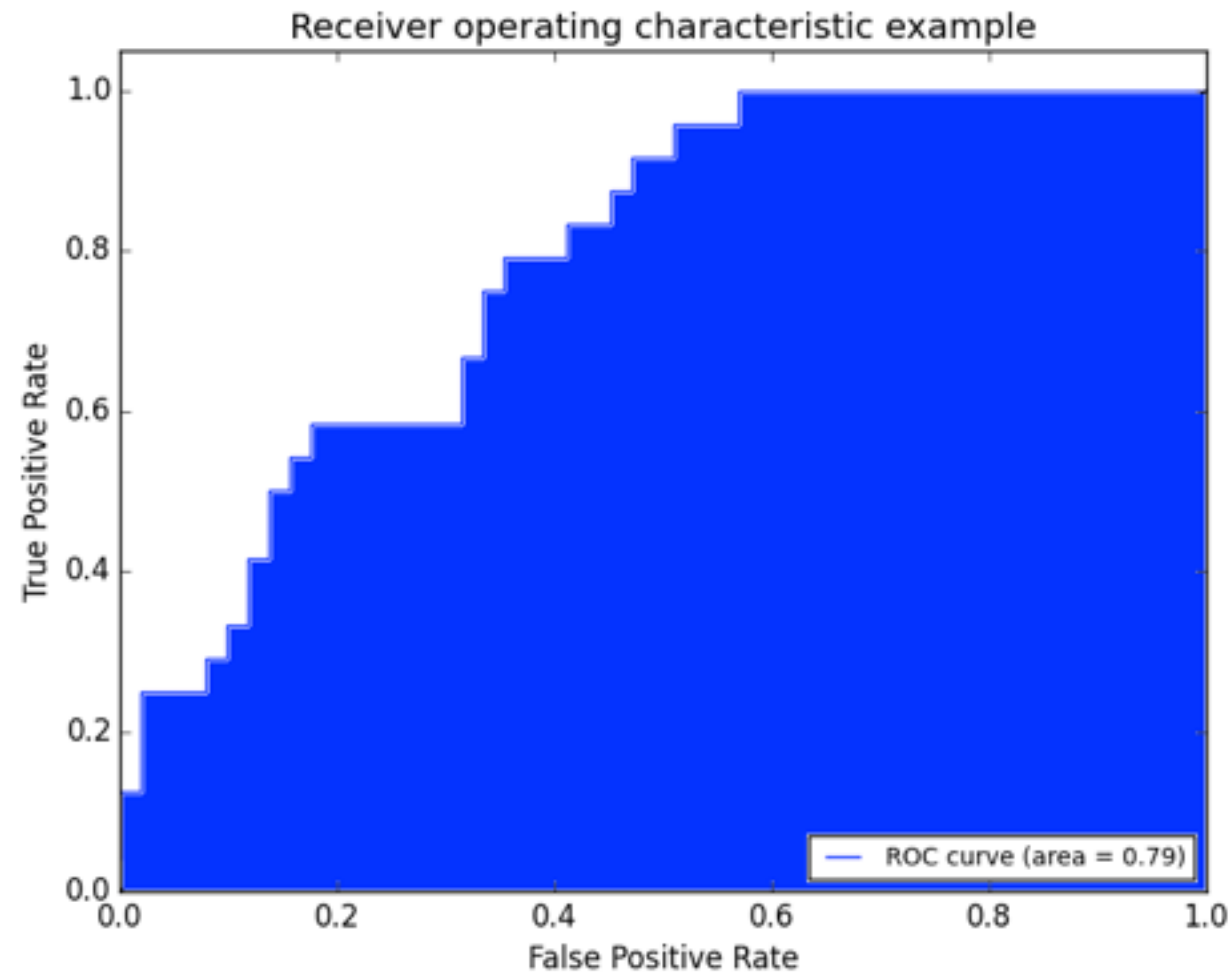
ADVANCED CLASSIFICATION METRICS

- Finally, we create a full curve that is described by TPR and FPR.



ADVANCED CLASSIFICATION METRICS

- With this curve, we can find the Area Under the Curve (AUC).



ADVANCED CLASSIFICATION METRICS

- ▶ If we have a TPR of 1 (all positives are marked positive) and FPR of 0 (all negatives are not marked positive), we'd have an AUC of 1. This means everything was accurately predicted.
- ▶ If we have a TPR of 0 (all positives are not marked positive) and an FPR of 1 (all negatives are marked positive), we'd have an AUC of 0. This means nothing was predicted accurately.
- ▶ An AUC of 0.5 would suggest randomness (somewhat) and is an excellent benchmark to use for comparing predictions (i.e. is my AUC above 0.5?).

ADVANCED CLASSIFICATION METRICS

- ▶ There are several other common metrics that are similar to TPR and FPR.

		<u>True class</u>			
		p	n		
<u>Hypothesized class</u>	Y	True Positives	False Positives	$fp\ rate = \frac{FP}{N}$	$tp\ rate = \frac{TP}{P}$
	N	False Negatives	True Negatives	$precision = \frac{TP}{TP+FP}$	$recall = \frac{TP}{P}$

- ▶ Sklearn has

Column totals:

P

N

$$F\text{-measure} = F1 = \frac{2TP}{2TP + FP + FN}$$

$$ACC = (TP + TN) / (TP + FP + FN + TN)$$

GUIDED PRACTICE

WHICH METRIC
SHOULD I USE?

ACTIVITY: WHICH METRIC SHOULD I USE?



EXERCISE

DIRECTIONS (15 minutes)

While AUC seems like a “golden standard”, it could be *further* improved depending upon your problem. There will be instances where error in positive or negative matches will be very important. For each of the following examples:

1. Write a confusion matrix: true positive, false positive, true negative, false negative. Then decide what each square represents for that specific example.
2. Define the *benefit* of a true positive and true negative.
3. Define the *cost* of a false positive and false negative.
4. Determine at what point does the cost of a failure outweigh the benefit of a success? This would help you decide how to optimize TPR, FPR, and AUC.

DELIVERABLE

Answers for each example

ACTIVITY: WHICH METRIC SHOULD I USE?



EXERCISE

DIRECTIONS (15 minutes)

Examples:

1. A test is developed for determining if a patient has cancer or not.
2. A newspaper company is targeting a marketing campaign for "at risk" users that may stop paying for the product soon.
3. You build a spam classifier for your email system.

DELIVERABLE

Answers for each example

INDEPENDENT PRACTICE

EVALUATING LOGISTIC REGRESSION WITH ALTERNATIVE METRICS

ACTIVITY: EVALUATING LOGISTIC REGRESSION



EXERCISE

DIRECTIONS (35 minutes)

[Kaggle's common online exercise](#) is exploring survival data from the Titanic.

1. Spend a few minutes determining which data would be most important to use in the prediction problem. You may need to create new features based on the data available. Consider using a feature selection aide in sklearn. For a worst case scenario, identify one or two strong features that would be useful to include in this model.

DELIVERABLE

Answers to the above question and a Logistic model on the Titanic data

ACTIVITY: EVALUATING LOGISTIC REGRESSION



EXERCISE

DIRECTIONS (35 minutes)

1. Spend 1-2 minutes considering which *metric* makes the most sense to optimize. Accuracy? FPR or TPR? AUC? Given the business problem of understanding survival rate aboard the Titanic, why should you use this metric?
1. Build a tuned Logistic model. Be prepared to explain your design (including regularization), metric, and feature set in predicting survival using any tools necessary (such as a fit chart). Use the starter code to get you going.

DELIVERABLE

Answers to the above question and a Logistic model on the Titanic data

OPENING

COMMUNICATING RESULT

WE BUILT A MODEL! NOW WHAT?

- ▶ We've built our model, but there is still a gap between your Notebook with plots/figures and a slideshow needed to present your results.
- ▶ Classes so far have focused on two core concepts:
 - ▶ developing consistent practices
 - ▶ interpreting metrics to evaluate and improve model performance
- ▶ But what does that mean to your audience?

WE BUILT A MODEL! NOW WHAT?

- ▶ Imagine how a non-technical audience might respond to the following statements:
 - ▶ The predictive model I built has an accuracy of 80%.
 - ▶ Logistic regression was optimized with L2 regularization.
 - ▶ Gender was more important than age in the predictive model because it has a larger coefficient.
 - ▶ Here's the AUC chart that shows how well the model did.

WE BUILT A MODEL! NOW WHAT?

- ▶ Who is your audience? Are they technical? What are their concerns?
- ▶ Remember: in a business setting, you may be *the only person* who can interpret what you've built.
- ▶ Some people may be familiar with basic visualization, but you will likely have to do a lot of “hand holding”.
- ▶ You need to be able to efficiently explain your results in a way that makes sense to **all** stakeholders (technical or not).

WE BUILT A MODEL! NOW WHAT?

- ▶ Today, we'll focus on communicating results for “simpler” problems, but this applies to any type of model you may work with.
- ▶ First, let's review classification metrics, review our knowledge, and talk about how we might communicate what we know.

REVIEW

BACK TO THE CONFUSION MATRIX

BACK TO THE CONFUSION MATRIX

- ▶ Confusion matrices allow for the interpretation of correct and incorrect predictions for *each class label*.
- ▶ It is the first step for the majority of classification metrics and goes deeper than just accuracy.

BACK TO THE CONFUSION MATRIX

- Let's recall our confusion matrix.

		<u>True class</u>			
		p	n		
<u>Hypothesized class</u>	Y	True Positives	False Positives	$\text{fp rate} = \frac{FP}{N}$	$\text{tp rate} = \frac{TP}{P}$
	N	False Negatives	True Negatives	$\text{precision} = \frac{TP}{TP+FP}$	$\text{recall} = \frac{TP}{P}$
Column totals:		P	N	$\text{accuracy} = \frac{TP+TN}{P+N}$	
				$\text{F-measure} = \frac{2}{1/\text{precision} + 1/\text{recall}}$	

ACTIVITY: KNOWLEDGE CHECK

ANSWER THE FOLLOWING QUESTIONS



EXERCISE

1. Without looking at the previous slide, how do we calculate the following?
 - a. Accuracy
 - b. True positive rate
 - c. False positive rate

DELIVERABLE

Answers to the above questions

INTRODUCTION

PRECISION AND RECALL

PRECISION AND RECALL

- ▶ Our previous metrics were primarily designed for less biased data problems: we could be interested in both outcomes, so it was important to generalize our approach.
- ▶ For example, we may be interested if a person will vote for a Republican or Democrat. This is a binary problem, but we're interested in both outcomes.

PRECISION AND RECALL

- ▶ Precision and recall, metrics built from the confusion matrix, focus on *information retrieval*, particularly when one class is more interesting than the other.
- ▶ For example, we may want to predict if a person will be a customer. We care much more about people who will be a customer of ours than people who won't.

PRECISION AND RECALL

- ▶ *Precision* aims to product a high amount of relevancy instead of irrelevancy.
- ▶ Precision asks, “Out of all of our positive predictions (both true positive and false positive), how many were correct?”
- ▶ *Recall* aims to see how well a model returns specific data (literally, checking whether the model can *recall* what a class label looked like).
- ▶ Recall asks, “Out of all of our positive class labels, how many were correct?”

ACTIVITY: KNOWLEDGE CHECK

ANSWER THE FOLLOWING QUESTIONS



EXERCISE

1. If the goal of the “recall” metric is to identify specific values of a class correctly, what other metric performs a similar calculation?

DELIVERABLE

Answers to the above question

THE MATH FOR RECALL

- ▶ Recall is the count of predicted *true positives* over the total count of that class label.
- ▶ This is the same as True Positive Rate or *sensitivity*.

		<u>True class</u>			
		p	n		
<u>Hypothesized class</u>	Y	True Positives	False Positives	$\text{fp rate} = \frac{FP}{N}$	$\text{tp rate} = \frac{TP}{P}$
	N	False Negatives	True Negatives	$\text{precision} = \frac{TP}{TP+FP}$	$\text{recall} = \frac{TP}{P}$
Column totals:		P	N	$\text{accuracy} = \frac{TP+TN}{P+N}$	
				$\text{F-measure} = \frac{2}{1/\text{precision} + 1/\text{recall}}$	

THE MATH FOR RECALL

- ▶ Imagine predicting the color of a marble as either red or green. There are 10 of each.
- ▶ If the model identifies 8 identifies 8 of the green marbles as green, the recall is $8 / 10 = 0.80$.
- ▶ However, this says nothing of the number of *red* marbles that are also identified as green.

THE MATH FOR PRECISION

- Precision, or positive predicted value, is calculated as the count of predicted true positives over the count of all values predicted to be positive.

		<u>True class</u>			
		p	n		
<u>Hypothesized class</u>	Y	True Positives	False Positives	$fp\ rate = \frac{FP}{N}$	$tp\ rate = \frac{TP}{P}$
	N	False Negatives	True Negatives	$precision = \frac{TP}{TP+FP}$	$recall = \frac{TP}{P}$
Column totals:		P	N	$accuracy = \frac{TP+TN}{P+N}$	
				$F\text{-measure} = \frac{2}{1/precision + 1/recall}$	

THE MATH FOR PRECISION

- ▶ Let's use our marble example again.
- ▶ If a model predicts 8 of the green marbles as green, then precision would be 1.00, because all marbles predicted as green were in fact green.
- ▶ Let's assume all red marbles were predicted correctly, and 2 green were predicted as red.
- ▶ The precision of red marbles would be $10 / (10 + 2) = 0.833$.

ANOTHER EXAMPLE

- ▶ For this example, we would have the following confusion matrix.

		True Class	
		Green	Red
Predicted Class	Green	8	4
	Red	12	12

- ▶ We could calculate precision for green marbles as $8 / (8 + 4) = 0.6666$.
- ▶ We could calculate recall for green marbles as $8 / (8 + 12) = 0.4000$.

ACTIVITY: KNOWLEDGE CHECK



ANSWER THE FOLLOWING QUESTIONS

- 1. What would the precision and recall be for the following confusion matrix (with “green” being “true”)?

	predicted_green	predicted_not_green
is_green	13	7
is_not_green	8	12

DELIVERABLE

Answers to the above question

THE DIFFERENCE BETWEEN PRECISION AND RECALL

- ▶ The key difference between the two is the attribution and value of error.
- ▶ Should our model be more pick in avoiding false positives (precision)?
- ▶ Or should it be more pick in avoiding false negatives (recall)?
- ▶ The answer should be determined by the problem you're trying to solve.

DEMO

UNDERSTANDING TRADEOFF

UNDERSTANDING TRADEOFF

- ▶ Let's consider the following data problem: we are given a data set in order to predict or identify traits for typically late flights.
- ▶ Optimizing toward recall, we could assume that every flight will be delayed.
- ▶ The trade-off, a lower precision, is that this could create even further delays, missed flights, etc.

UNDERSTANDING TRADEOFF

- ▶ Optimizing toward precision, we would specifically look to identify flights that will be late.
- ▶ The trade-off here would be lower recall. We might miss flights that would be delayed, causing a strain on the system.

UNDERSTANDING TRADEOFF

- ▶ This plot is based on choosing decision line thresholds, much like the AUC figure from the previous class.
- ▶ In terms of modeling delays, this would be like moving the decision line for lateness from a probability of 0.01 up to 0.99, and then calculating the precision and recall at each decision.

UNDERSTANDING TRADEOFF

- ▶ Interpreting our plot, there's a few interesting nuggets compared to the benchmark (blue line):
 - ▶ At a lower recall (below 0.2), there is a noticeable lower precision in the model.
 - ▶ Beyond 0.2 recall, the model outperforms the benchmark.
- ▶ Whether we're optimizing for recall or precision, this plot helps us decide based on the 0.3 threshold.

GUIDED PRACTICE

COST BENEFIT ANALYSIS

ACTIVITY: COST BENEFIT ANALYSIS

DIRECTIONS (15 minutes)

One tool that complements the confusion matrix is cost-benefit analysis, where you attach a *value* to correctly and incorrectly predicted data.

Like the Precision-Recall trade off, there is a balancing point to the *probabilities* of a given position in the confusion matrix, and the *cost* or *benefit* to that position. This approach allows you to not only add a weighting system to your confusion matrix, but also to speak the language of your business stakeholders (i.e. communicate your values in dollars!).



EXERCISE

ACTIVITY: COST BENEFIT ANALYSIS



EXERCISE

DIRECTIONS

Consider the following marketing problem:

As a data scientist working on marketing spend, you've build a model that reduces user churn--the number of users who decide to stop paying for a product--through a marketing campaign. Your model generates a confusion matrix with the following probabilities (these probabilities are calculated as the value in that position over the sum of the sample):

	TP: 0.2		FP: 0.2	

	FN: 0.1		TN: 0.5	

ACTIVITY: COST BENEFIT ANALYSIS



EXERCISE

DIRECTIONS (15 minutes)

In this case:

- The *benefit* of a true positive is the retention of a user (\$10 for the month)
- The *cost* of a false positive is the spend of the campaign per user (\$0.05)
- The *cost* of a false negative (someone who could have retained if sent the campaign) is, effectively, 0 (we didn't send it... but we certainly didn't benefit!)
- The *benefit* of a true negative is 0: No spend on users who would have never retained.

To calculate Cost-Benefit, we'll use this following function:

$$(P(TP) * B(TP)) + (P(TN) * B(TN)) + (P(FP) * C(FP)) + (C(FN) * C(FN))$$

which for our marketing problem, comes out to this:

$$(.2 * 10) + (.5 * 0) - (.2 * .05) - (.1 * 0)$$

or \$1.99 per user targeted.

ACTIVITY: COST BENEFIT ANALYSIS



EXERCISE

FOLLOW UP QUESTIONS

Think about precision, recall, and cost benefit analysis to answer the following questions:

1. How would you rephrase the business problem if your model was optimizing toward *precision*? i.e., How might the model behave differently, and what effect would it have?
2. How would you rephrase the business problem if your model was optimizing toward *recall*?
3. What would the most ideal model look like in this case?

DELIVERABLE

Answers to the above questions

INTRODUCTION

SHOWING WORK

SHOWING WORK

- ▶ We've spent a lot of time exploring our data and building a reasonable model that performs well.
- ▶ However, if we look at our visuals, they are most likely:
 - ▶ Statistically heavy: Most people don't understand histograms.
 - ▶ Overly complicated: Scatter matrices produce too much information.
 - ▶ Poorly labeled: Code doesn't require adding labels, so you may not have added them.

SHOWING WORK

- ▶ In order to convey important information to our audience, make sure our charts are:
 - ▶ Simplified
 - ▶ Easily interpretable
 - ▶ Clearly labeled

SIMPLIFIED

- ▶ At most, you'll want to include figures that either explain a variable on its own or explain that variable's relationship with a target.
- ▶ If your model used a data transformation (like natural log), just visualize the original data.
- ▶ Try to remove any unnecessary complexity.

EASILY INTERPRETABLE

- ▶ Any stakeholder looking at a figure should be seeing the exact same thing you're seeing.
- ▶ A good test for this is to share the visual with others less familiar with the data and see if they come to the same conclusion.
- ▶ How long did it take them?

CLEARLY LABELED

- ▶ Take the time to clearly label your axis, title your plot, and double check your scales - especially if the figures should be comparable.
- ▶ If you're showing two graphs side by side, they should follow the same Y axis.

QUESTION TO ASK

- ▶ When building visuals for another audience, ask yourself these questions:
 - ▶ **Who:** Who is my target audience for the visual?
 - ▶ **What:** What do they already know about this project? What do they need to know?
 - ▶ **How:** How does my project affect this audience? How might they interpret (or misinterpret) the data?

DEMO

VISUALIZING MODELS OVER VARIABLES

VISUALIZING MODELS OVER VARIABLES

- ▶ One effective way to explain your model over particular variables is to plot the predicted values against the most explanatory variables.
- ▶ For example, in logistic regression, plotting the probability of a class against a variable can help explain the range of effect of the model.

VISUALIZING MODELS OVER VARIABLES

- ▶ We'll use the flight delay data for all following examples. Let's build our first model and plot.
- ▶ Open the starter code from the class repo and follow along.

VISUALIZING MODELS OVER VARIABLES

```
# read in the file and generate a quick model (assume we've done the
data exploration already)
import pandas as pd
import sklearn.linear_model as lm
import matplotlib.pyplot as plt

df = pd.read_csv('../assets/dataset/flight_delays.csv')

df = df.join(pd.get_dummies(df[ 'DAY_OF_WEEK' ], prefix='dow' ))
df = df[df.DEP_DEL15.notnull()].copy()
```

VISUALIZING MODELS OVER VARIABLES

```
# Build a model
model = lm.LogisticRegression()
features = ['dow_1', 'dow_2', 'dow_3', 'dow_4', 'dow_5', 'dow_6']
model.fit(df[features + ['CRS_DEP_TIME']], df['DEP_DEL15'])

df['probability'] = model.predict_proba(df[features +
['CRS_DEP_TIME']]).T[1]
```

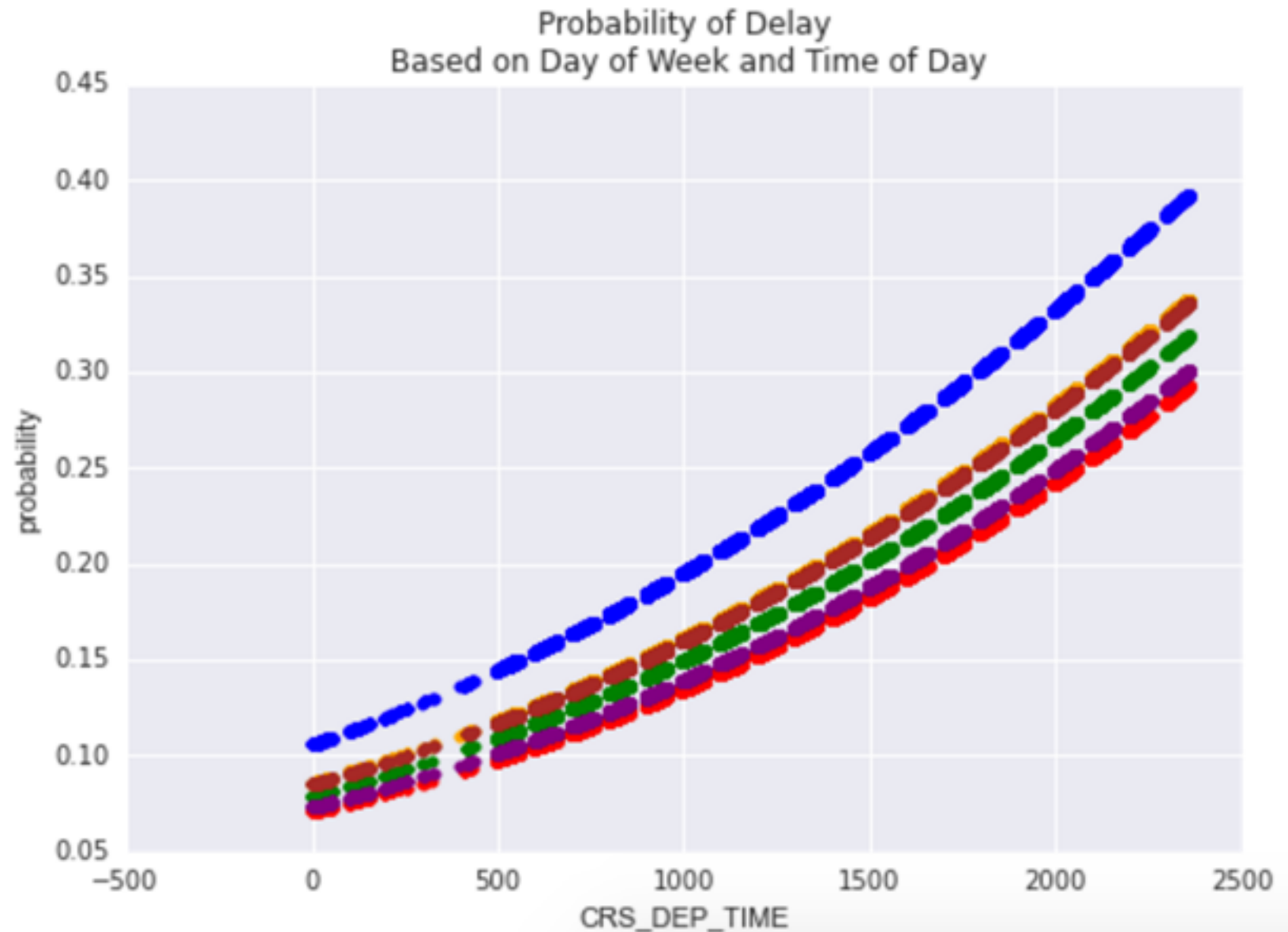
VISUALIZING MODELS OVER VARIABLES

```
# Create a plot
ax = plt.subplot(111)
colors = ['blue', 'green', 'red', 'purple', 'orange', 'brown']
for e, c in enumerate(colors):
    df[df[features[e]] == 1].plot(x='CRS_DEP_TIME', y='probability',
kind='scatter', color = c, ax=ax)

ax.set(title='Probability of Delay\n Based on Day of Week and Time of
Day')
```

VISUALIZING MODELS OVER VARIABLES

- ▶ This visual can help showcase the range of effect on delays from both day of the week and time of day.
- ▶ Given this model, some days are more likely to have delays than others.
- ▶ The likelihood of delay increases as the day goes on.



ACTIVITY: TRY IT OUT



EXERCISE

DIRECTIONS

1. Adjust the model to make delay predictions using airlines instead of day of week, and time, then plot the effect on CRS_DEP_TIME=1.
1. Try plotting the inverse: pick either model and plot the effect on CRS_DEP_TIME=0.

DELIVERABLE

The new plots

DEMO

VISUALIZING PERFORMANCE AGAINST BASELINE

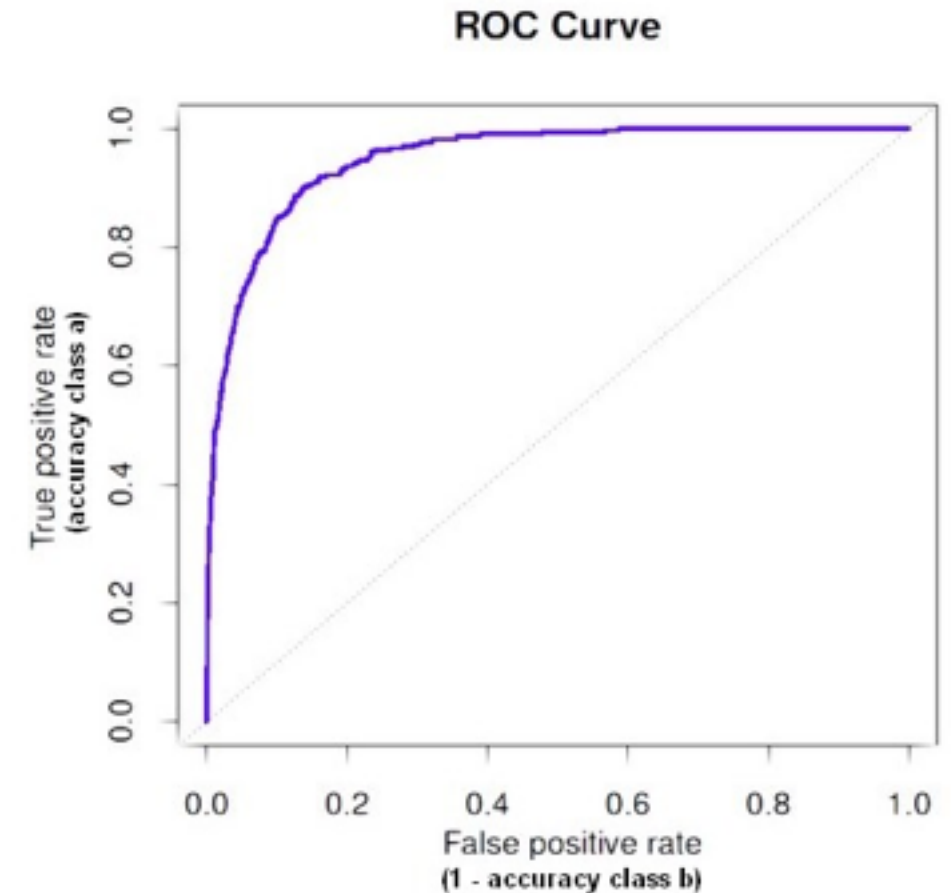
VISUALIZING PERFORMANCE AGAINST BASELINE

- ▶ Another approach of visualization is the effect of your model against a baseline, or - even better - against previous models.
- ▶ Plots like this will also be useful when talking to your peers - other data scientists or analysts who are familiar with your project and interested in the progress you've made.

VISUALIZING PERFORMANCE AGAINST BASELINE

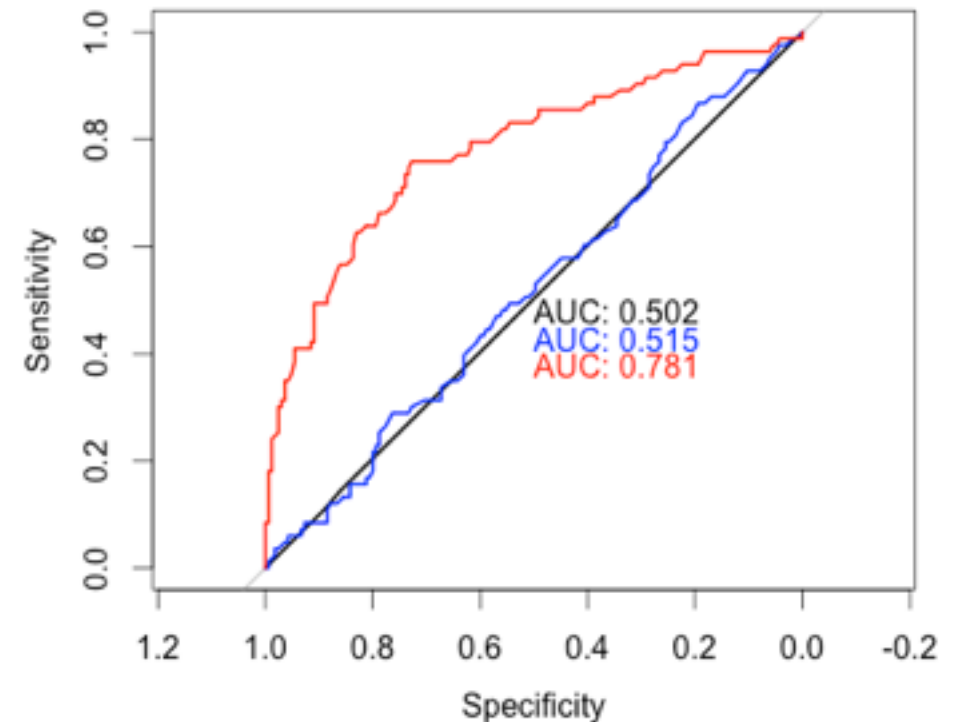
► For classification, we've practiced plotting AUC and precision-recall plots. Consider the premise of each:

- AUC plots explain and represent “accuracy” as having the largest area under the curve. Good models will be high and to the left.
- For precision-recall plots, it will depend on the *cost* requirements. Either a model will have good recall at the cost of precision or vice versa.



VISUALIZING PERFORMANCE AGAINST BASELINE

- ▶ When comparing multiple models:
 - ▶ For AUC plots, you'll be interested in which model has the *largest* area under the curve.
 - ▶ For precision-recall plots, based on the cost requirement, you are looking at which model has the best precision given the same recall, or the best recall given the same precision.



VISUALIZING PERFORMANCE AGAINST BASELINE

- ▶ Follow along with the starter code located in the class repo.
- ▶ We've plotted several models for AUC: a dummy model and additional features.

```
model0 = dummy.DummyClassifier()  
model0.fit(df[features[1:-1]], df.DEP_DEL15)  
df['probability_0'] = model0.predict_proba(df[features[1:-1]]).T[1]
```

```
model = lm.LogisticRegression()  
model.fit(df[features[1:-1]], df.DEP_DEL15)  
df['probability_1'] = model.predict_proba(df[features[1:-1]]).T[1]
```

VISUALIZING PERFORMANCE AGAINST BASELINE

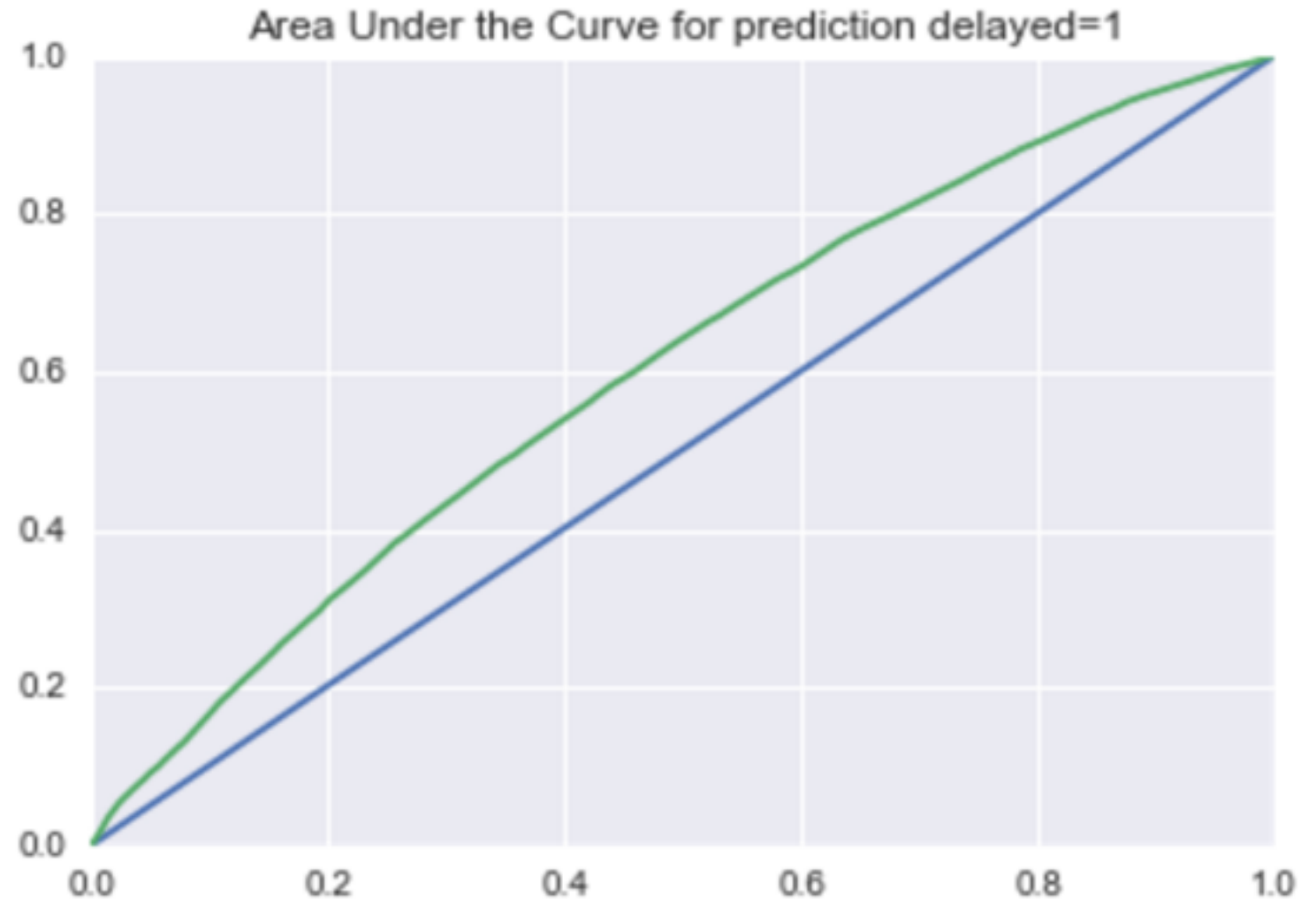
```
ax = plt.subplot(111)
vals = metrics.roc_curve(df.DEP_DEL15, df.probability_0)
ax.plot(vals[0], vals[1])
vals = metrics.roc_curve(df.DEP_DEL15, df.probability_1)
ax.plot(vals[0], vals[1])

ax.set(title='Area Under the Curve for prediction delayed=1',
       ylabel='TRP', xlabel='FRP', xlim=(0, 1), ylim=(0, 1))
```

VISUALIZING PERFORMANCE AGAINST BASELINE

► This plot showcases:

1. The model using data outperforms a baseline dummy model.
1. By adding other features, there's some give and take with probability as the model gets more complicated.



ACTIVITY: TRY IT OUT



EXERCISE

DIRECTIONS

1. In a similar approach, use the sklearn `precision_recall_curve` function to enable you to plot the precision-recall curve of the four models from above. Keep in mind precision in the first array is returned from the function, but the plot shows it as the y-axis.
2. Explain what is occurring when the recall is below 0.2.
3. Based on this performance, is there a clear winner at different thresholds?
4. **Bonus:** Redo both the AUC and precision-recall curves using models that have been cross validated using `kfold`. How do these new figures change your expectations for performance?

DELIVERABLE

The new plots and associated answers

INDEPENDENT PRACTICE

PROJECT PRACTICE

ACTIVITY: PROJECT PRACTICE



EXERCISE

DIRECTIONS (45 minutes)

Using models built from the flight data problem earlier in class, work through the same problems. Your data and models should already be accessible. Your goals:

1. There are *many* ways to manipulate this data set. Consider what is a proper "categorical" variable, and keep *only* what is significant. You will easily have 20+ variables. Aim to have at least three visuals that clearly explain the relationship of variables you've used against the predictive survival value.
2. Generate the AUC or precision-recall curve (based on which you think makes more sense), and have a statement that defines, compared to a baseline, how your model performs and any caveats. For example: "My model on average performs at x rate, but the features under-perform and explain less of the data at these thresholds." Consider this as practice for your own project, since the steps you'll take to present your work will be relatively similar.

DELIVERABLE

New models and performance statement

CONCLUSION

TOPIC REVIEW

REVIEW AND NEXT STEPS

- ▶ What do precision and recall mean? How are they similar and different to True Positive Rate and False Positive Rate?
- ▶ How does cost benefit analysis play a role in building models?
- ▶ What are at least two very important details to consider when creating visuals for a project's stakeholders?
- ▶ Why would an AUC plot work well for a data science audience but not for a business audience? What would be a more effective visualization for that group?

COURSE

**BEFORE NEXT
CLASS**

BEFORE NEXT CLASS

UPCOMING

► Project: Unit Project 4

LESSON

CREDITS

THANKS FOR THE FOLLOWING

CITATIONS

- ▶ Title, Author: link
- ▶ Title, Author: link
- ▶ Title, Author: link

LESSON

Q & A

LESSON

EXIT TICKET

**DON'T FORGET TO FILL OUT YOUR EXIT
TICKET**

THANKS!

NAME

- Optional Information:
- Email?
- Website?
- Twitter?