

Advanced Metrics and Communicating Results

Ivan Corneillet

Data Scientist

Learning Objectives

After this lesson, you should be able to:

- Evaluate a model using advanced metrics such as confusion matrix and ROC/AUC curves
- Explain the trade-offs between false positives vs. false negatives
- Describe the difference between visualization for presentations vs. exploratory data analysis
- Identify the components of a concise and convincing report and how they relate to specific audiences/stakeholders



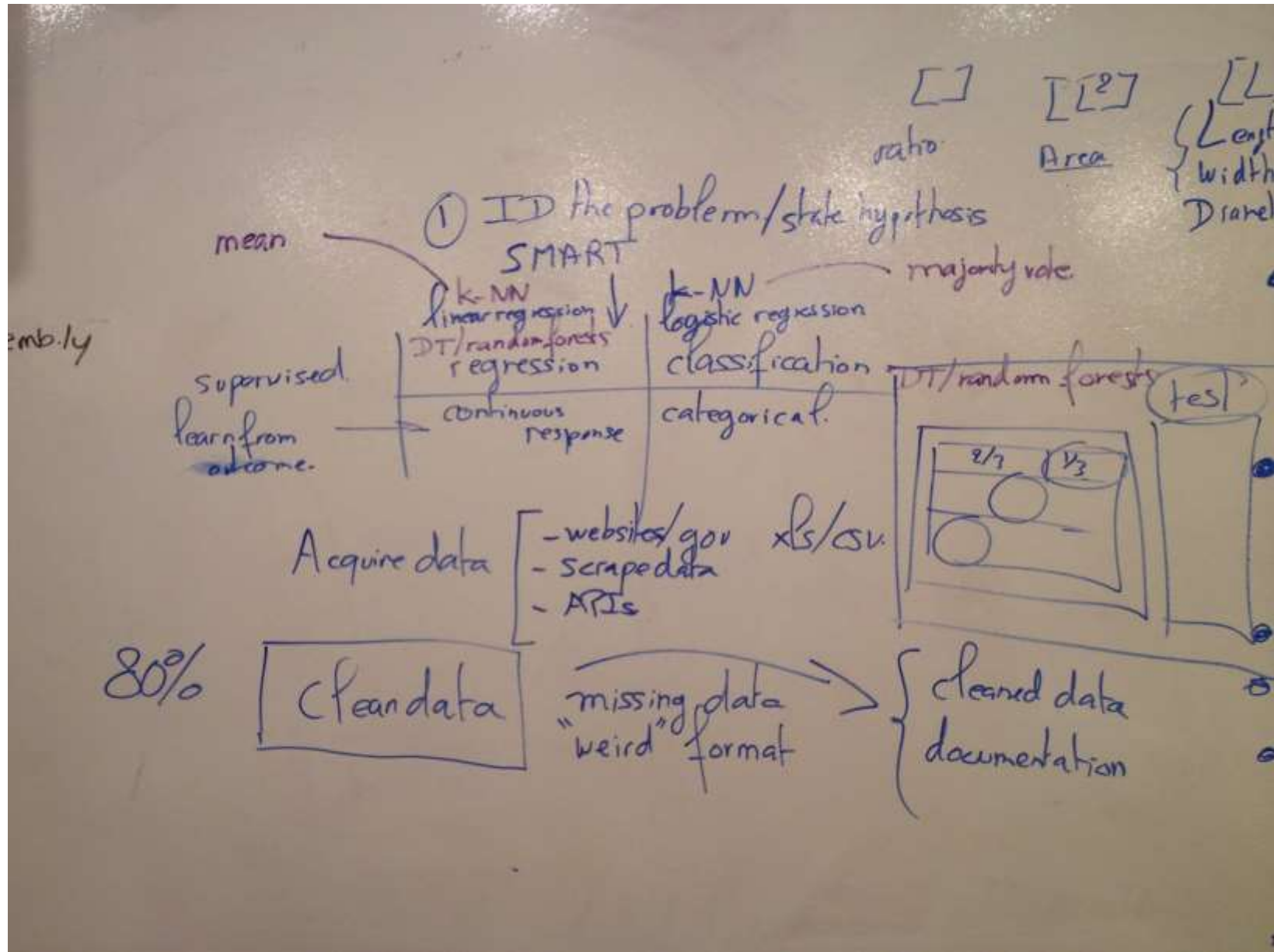
DS

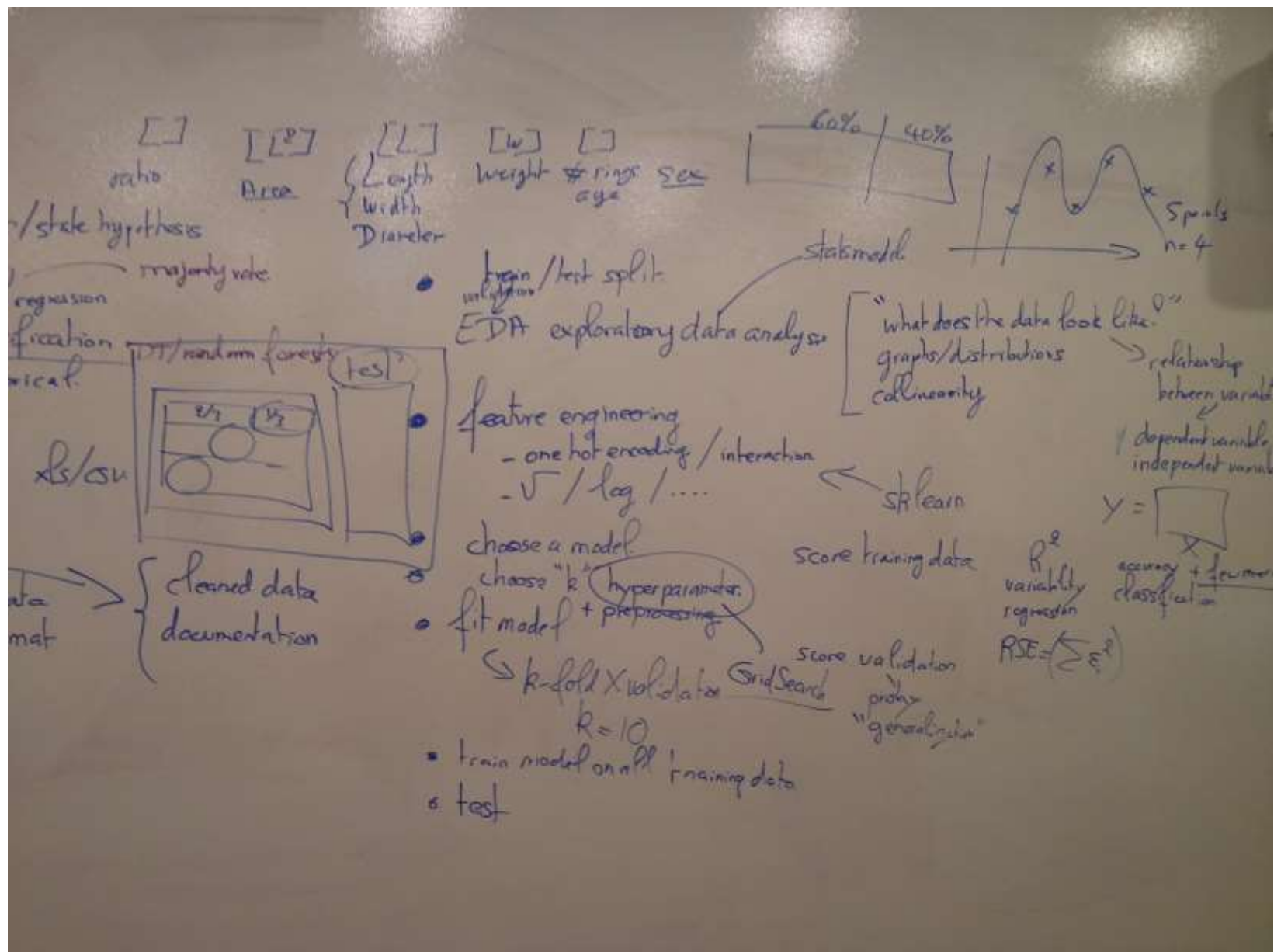
Announcements and Exit Tickets

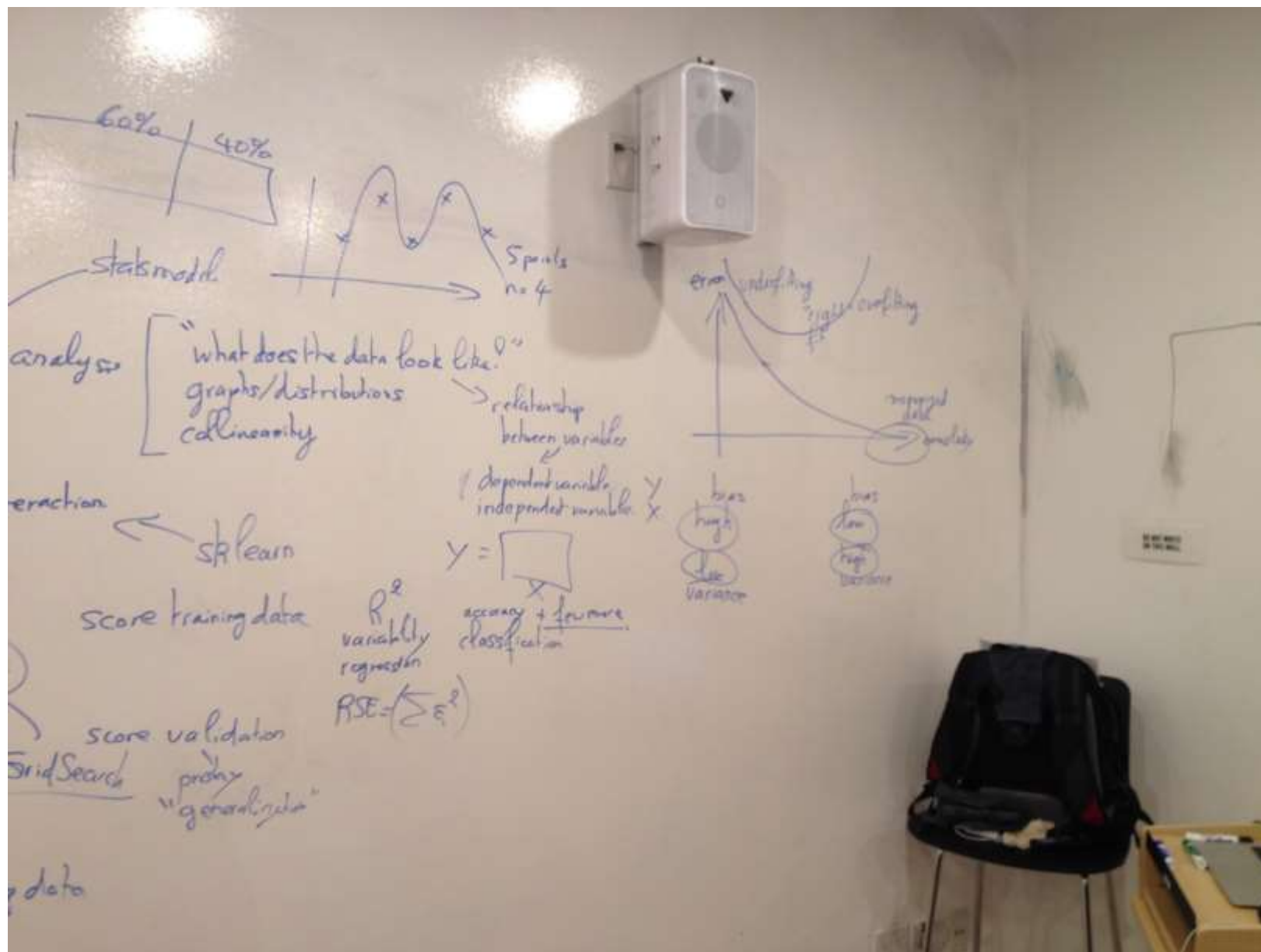


DS

Review







A black circle containing the white text "DS".

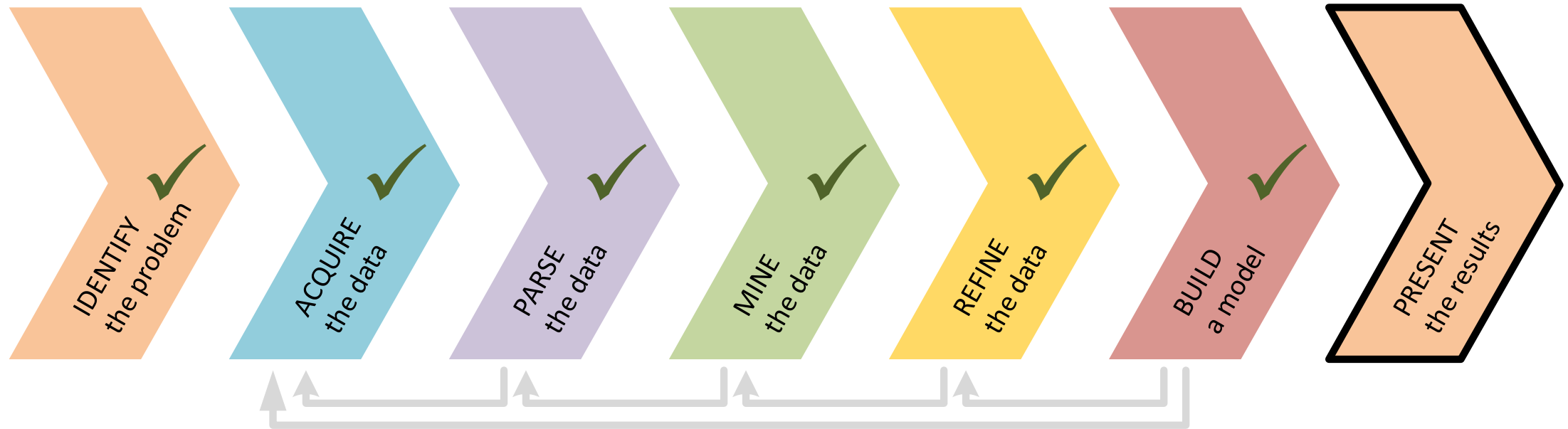
DS

Today

Today, we are wrapping Unit 2 – Foundation of Modeling

Research Design and Data Analysis	Research Design	Data Visualization in <i>pandas</i>	Statistics	Exploratory Data Analysis in <i>pandas</i>
Foundations of Modeling	Linear Regression (sessions 6, 7, and 10)	Classification Models (k-Nearest Neighbors, Logistic Regression) (sessions 8, 9, and 10)	Evaluating Model Fit (sessions 5, 6, 7, and 11)	Presenting Insights from Data Models (session 11)
Data Science in the Real World	Decision Trees and Random Forests	Time Series Models	Natural Language Processing	Databases

... as well as the first full pass of the Data Science Workflow



Here's what's happening today:

- Announcements and Exit Tickets
- Review
- **⑥ Build a Model | Advanced metrics**
 - Confusion Matrix
 - True Positive and False Positive Rates, ROC, and AUC
 - Plotting the ROC/AUC
- Codealong for ROC/AUC
- **⑦ Present the Results | Communicating Results**
 - Showing our Work
 - Codealong to pretty up graphs
- Review
- Exit Tickets



DS

Pre-Work

Pre-Work

Before this lesson, you should already be able to:

- Explain the concepts of cross-validation, logistic regression, and overfitting
- Know how to build and evaluate some classification models in *sklearn* using cross-validation

DS

⑥ Build a Model

Advanced Metrics

Advanced Metrics

- Accuracy is only one of several metrics used when solving for a classification problem
 - E.g., if we know a prediction is 75% accurate, accuracy doesn't provide any insight into why the 25% was wrong. Was it wrong *equally* across all class labels? Did it just guess one class label for all predictions and 25% of the data was just the other label?
- It's important to look at other metrics to fully understand the problem

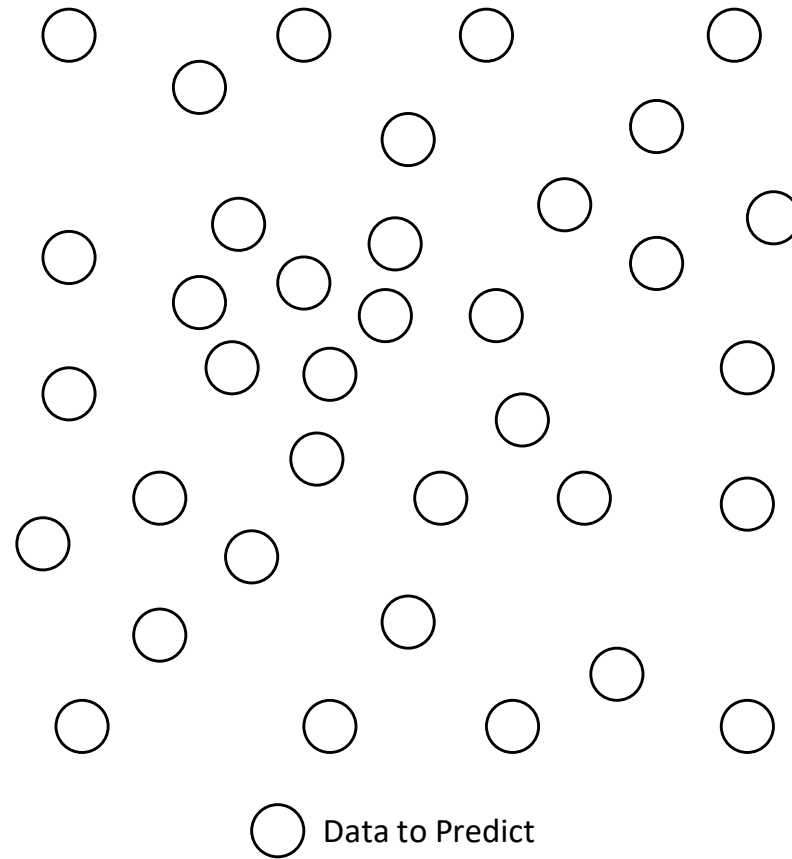
- Accuracy
 - How many observations that we predicted were correct? This is a value we'd want to increase (like R^2)
- Misclassification rate
 - Directly opposite of accuracy
 - Of all the observations we predicted, how many were incorrect? This is a value we'd want to decrease (like the mean squared error)

DS

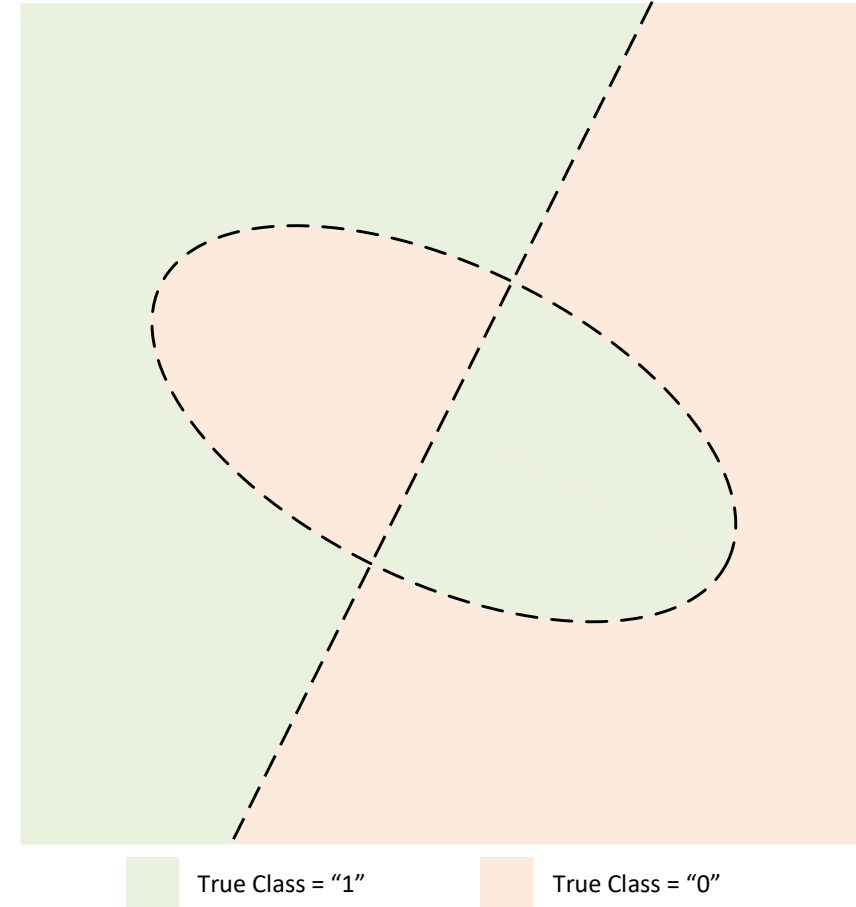
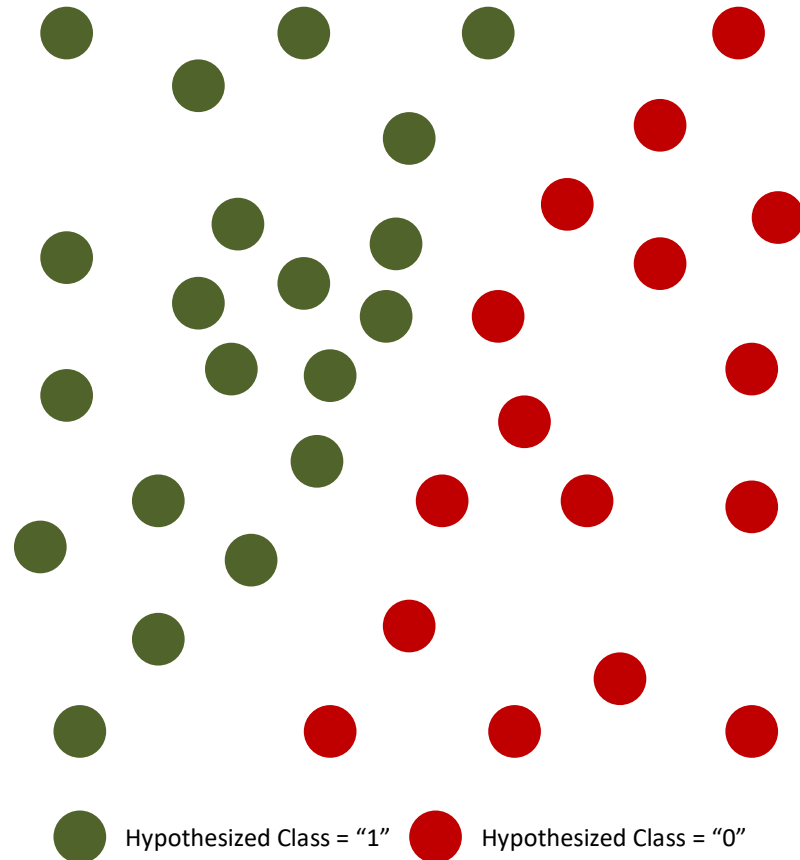
⑥ Build a Model

Confusion Matrix

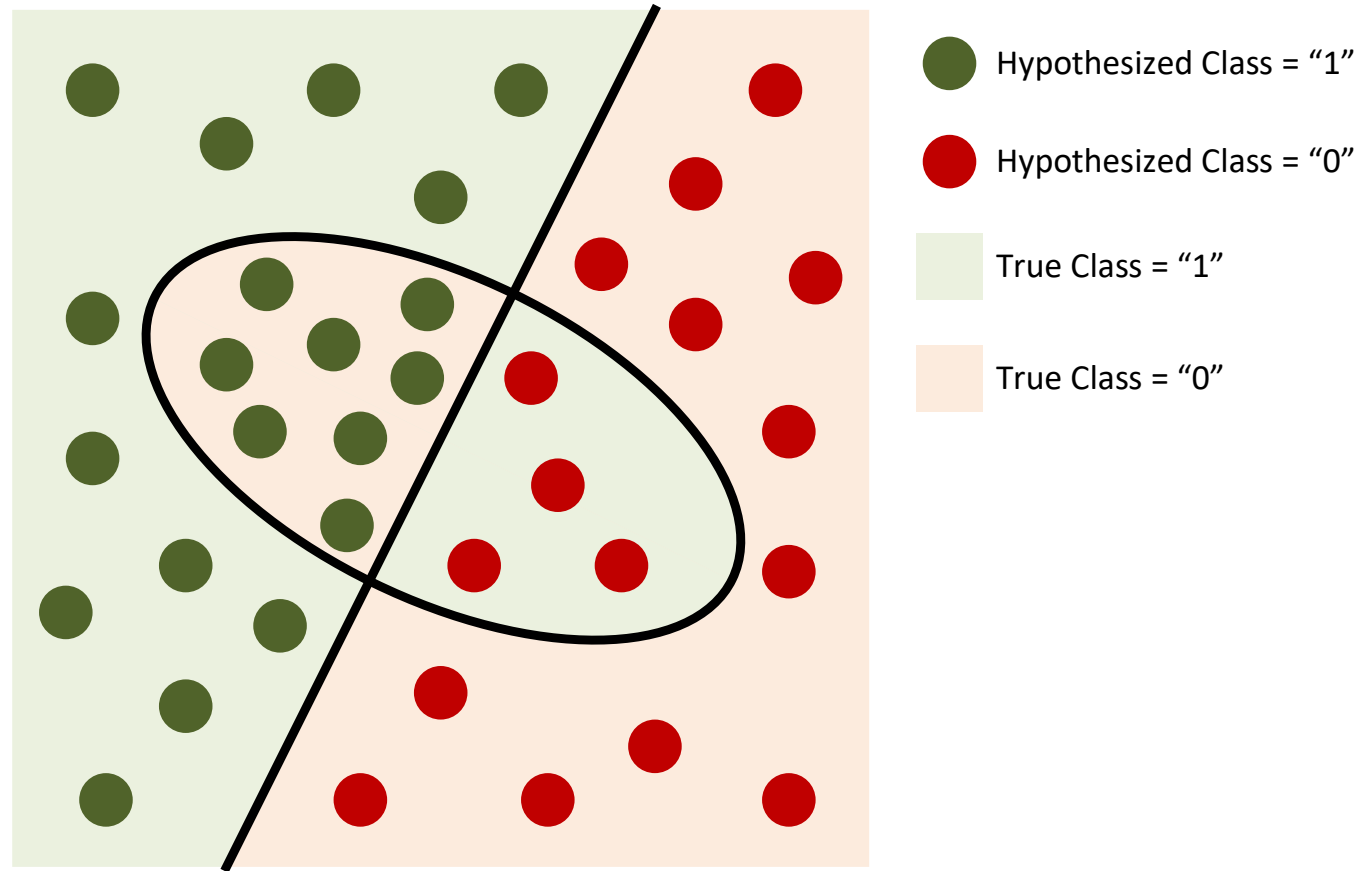
Stepping back | Let's say we want to classify this data:



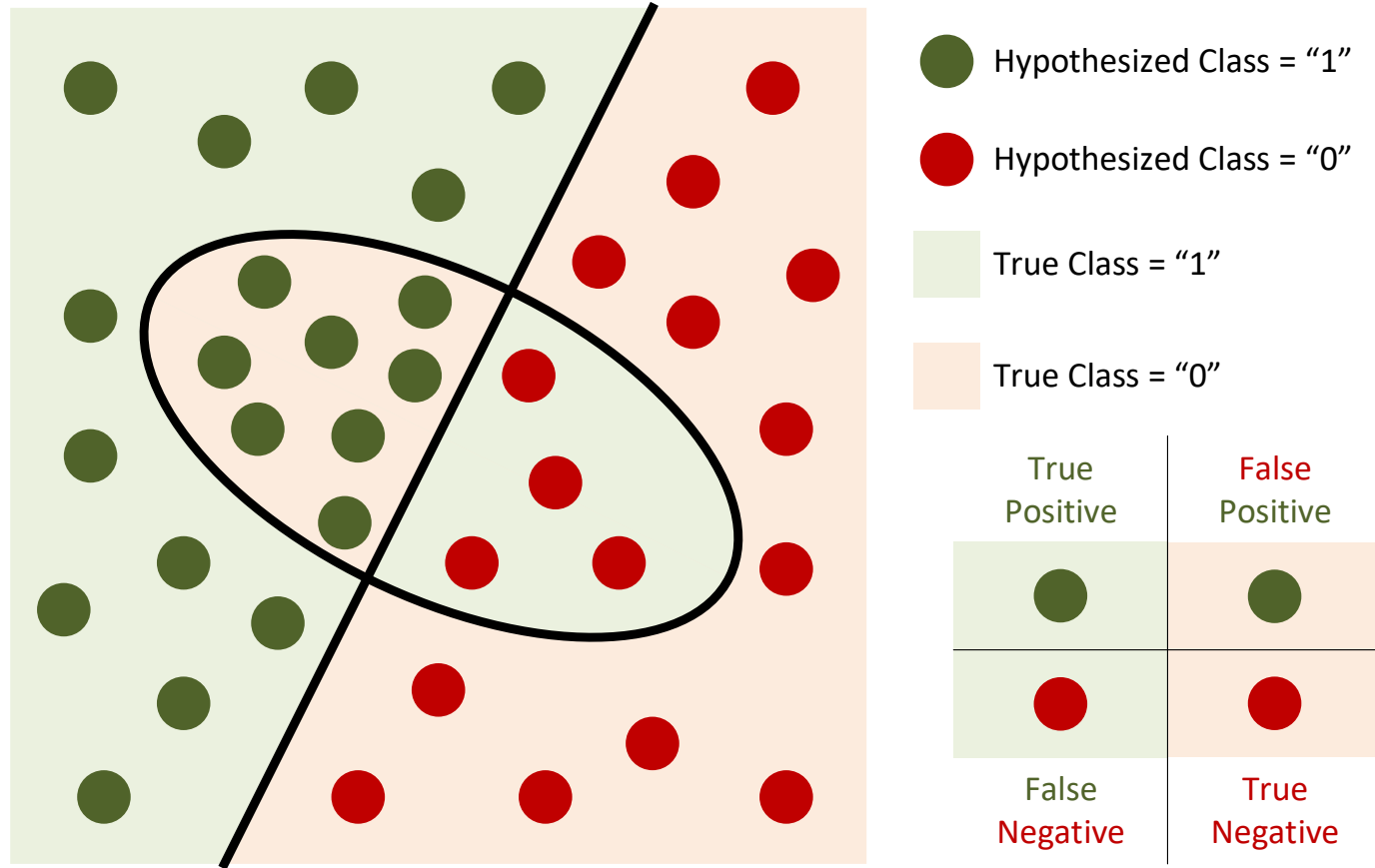
Hypothesized and true classes don't necessarily match



Putting hypothesized and true classes together, we get 4 possibilities



We can rearrange these 4 possibilities into a 2x2 table



Confusion Matrix (a.k.a., Contingency Table or Error Matrix)

		True Class	
		1	0
Hypothesized Class	1 ●	True Positives (TP) ●	False Positives (FP) <i>(type I error)</i> ●
	0 ■	False Negatives (FN) <i>(type II error)</i> ■	True Negatives (TN) ■
Total Columns		$P = TP + FN$	$N = FP + TN$

- A confusion matrix is a specific table layout that allows visualization of the performance of a supervised learning algorithm
- Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class
- The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e., commonly mislabeling one as another)

DS

⑥ Build a Model

Codealong – Part A
Confusion Matrix

DS

⑥ Build a Model

Activity / Interpreting the confusion matrix

Activity | Interpreting the confusion matrix



EXERCISE

DIRECTIONS (10 minutes)

1. Use the variables defined in the confusion matrix (TP , FN , FP , TN , P , and N) to calculate the answers to the following questions:
 - a. Overall, how often is the classifier correct?
 - b. When the classifier predicts yes, how often is it correct?
 - c. How often does the yes condition actually occur in our sample?
 - d. When it's actually yes, how often does the classifier predict yes?
 - e. When it's actually no, how often does the classifier predict yes?
 - f. When it's actually no, how often does it predict no?
 - g. Overall, how often is the classifier wrong?

Activity | Interpreting the confusion matrix (cont.)



EXERCISE

DIRECTIONS (cont.)





2. Given a medical exam that tests for cancer ($1 = \text{Cancer}$, $0 = \text{Cancer free}$), use the variables defined in the confusion matrix (TP , FN , FP , TN , P , and N) to calculate the answers to the following questions:
 - a. How often is it correct when it identify patients with cancer?
 - b. How often does it correctly identify patients without cancer?
 - c. How often does it trigger a “false alarm” by saying a patient has cancer when they actually don’t?
 - d. How often does it correctly identify patients with cancer?
3. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions





$\frac{TP+TN}{P+N}$				$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FP}$
$\frac{P}{P+N}$	$\frac{P}{P+N}$	$\frac{P}{P+N}$		$\frac{TP}{P}$	$\frac{TP}{P}$	$\frac{TP}{P}$	$\frac{TP}{P}$
$\frac{FP}{N}$	$\frac{FP}{N}$	$\frac{FP}{N}$	$\frac{FP}{N}$	$\frac{TN}{N}$	$\frac{TN}{N}$	$\frac{TN}{N}$	$\frac{TN}{N}$
$\frac{FN+FP}{P+N}$	$\frac{FP+FN}{P+N}$	$\frac{FN+FP}{P+N}$	$\frac{FP+FN}{P+N}$				

Activity | Interpreting the confusion matrix (cont.)

		True Class			
		1	0		
Hypothesized Class	1	 True Positives (TP)	 False Positives (FP) <i>(type I error)</i>	Question: Overall, how often is the classifier correct? Answer: $\frac{TP+TN}{P+N}$ <i>(accuracy)</i>	When the classifier predicts yes, how often is it correct? Answer: $\frac{TP}{TP+FP}$ <i>(precision)</i>
	0	 False Negatives (FN) <i>(type II error)</i>	 True Negatives (TN)	How often does the yes condition actually occur in our sample? Answer: $\frac{P}{P+N}$ <i>(prevalence)</i>	When it's actually yes, how often does the classifier predict yes? Answer: $\frac{TP}{P}$ <i>(TPR, sensitivity, recall)</i>
Total Columns		$P = TP + FN$	$N = FP + TN$	When it's actually no, how often does the classifier predict yes? Answer: $\frac{FP}{N}$ <i>(FPR, fall-out)</i>	When it's actually no, how often does it predict no? Answer: $\frac{TN}{N}$ <i>(specificity)</i>
				Overall, how often is the classifier wrong? Answer: $\frac{FP+FN}{P+N}$ <i>(misclassification rate)</i>	

Linear regression DT/random forest regression	Logistic class				
$\frac{TP}{TP+FP}$ $\frac{TP}{TP+FP}$	$\frac{TP}{TP+FP}$ $\frac{TP}{P}$	$\frac{TN}{N}$	$\frac{TN}{TN+FN}$ $\frac{TN}{N}$	$\frac{TN}{N}$	
$\frac{FP}{TP+FP}$ $\frac{FP}{N}$	$\frac{FP}{N}$ $\frac{P}{P+N}$	$\frac{TP}{P}$	$\frac{TP}{P}$?	
P	P		$\frac{TP}{P}$		

Activity | Interpreting the confusion matrix (cont.)

		True Class			
		Has Cancer	Doesn't have cancer		
Hypothesized Class	Predict Cancer ●	 True Positives (TP)	 False Positives (FP) <i>(type I error)</i>	<i>How often is it correct when it identify patients with cancer?</i> Answer: $\frac{TP}{TP+FP}$ <i>(precision)</i>	<i>How often does it correctly identify patients without cancer?</i> Answer: $\frac{TN}{N}$ <i>(specificity)</i>
	Predict No Cancer ■	 False Negatives (FN) <i>(type II error)</i>	 True Negatives (TN)	<i>How often does it trigger a "false alarm" by saying a patient has cancer when they actually don't?</i> Answer: $\frac{FP}{N}$ <i>(FPR, fall-out)</i>	<i>How often does it correctly identify patients with cancer?</i> Answer: $\frac{TP}{P}$ <i>(TPR, sensitivity, recall)</i>
Total Columns		$P = TP + FN$	$N = FP + TN$	<i>How many patients have cancer?</i> Answer: $\frac{P}{P+N}$ <i>(prevalence)</i>	

DS

⑥ Build a Model

Activity / Interpreting the confusion matrix – Take 2

Activity | Interpreting the confusion matrix – Take 2



EXERCISE

DIRECTIONS (5 minutes)

1. We trained a binary classifier and got the following hypothesized probabilities (\hat{p}) for the samples in the table.
 - a. What are the hypothesized classes (\hat{c})?
 - b. Are the samples true/false positive/negative?
2. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions

Activity | Interpreting the confusion matrix – Take 2 (cont.)





#	$\hat{p} = P(c = 1)$	\hat{c}	c	<i>True/False Positive/Negative</i>
1	.44	0	1	FN
2	.29	0	0	TN
3	.98	1	1	TP
4	.69	1	0	FP
5	.07	0	1	FN

DS

⑥ Build a Model

True and False Positive Rates, ROC, and AUC

True Positive Rate, $TPR = \frac{TP}{P}$

		True Class	
		1	0
Hypothesized Class	1	 True Positives (TP)	 False Positives (FP) (type I error)
	0	 False Negatives (FN) (type II error)	 True Negatives (TN)
Total Columns		$P = TP + FN$	$N = FP + TN$





▸ When it's actually yes, how often does the classifier predict yes?

▸ A.k.a., "Sensitivity"

▸ E.g., given a medical exam that tests for cancer, how often does it correctly identify patients with cancer?

▸ Likewise, this can be inverted: how often does a test *correctly* identify patients without cancer

False Positive Rate, $FPR = \frac{FP}{N}$

		True Class	
		1	0
Hypothesized Class	1	 True Positives (TP)	 False Positives (FP) <i>(type I error)</i>
	0	 False Negatives (FN) <i>(type II error)</i>	 True Negatives (TN)
Total Columns		$P = TP + FN$	$N = FP + TN$

‣ When it's actually no, how often does the classifier predict yes?

‣ A.k.a., “Fall-out”

‣ E.g., given a medical exam that tests for cancer, how often does it trigger a “false alarm” by saying a patient has cancer when they actually don't?

‣ Likewise, this can be also inverted: how often does a test *incorrectly* identify patients as being cancer-free when they might actually have cancer!

True positive and false positive rates

- We can split up the accuracy of each label by using true positive and false positive rates. Using them, we can get a much clearer picture of where predictions begin to fall apart
- A good classifier would have a true positive rate approaching 1, and a false positive rate approaching 0. In a binary problem (say, predicting if someone smokes or not), it would accurately predict all of the smokers as smokers, and not accidentally predict any of the non-smokers as smokers

DS

⑥ Build a Model

Activity / Introduction to the ROC space

Activity | Introduction to the ROC space



EXERCISE

DIRECTIONS (5 minutes)

1. Calculate TPR and FPR for the four confusion matrices in the handout and place them in the ROC space (TPR as a function of FPR)
2. How would you classify these four cases as a function of their performance (e.g., better or worse)
3. What does the ROC space tells you?
4. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions

Activity | Introduction to the ROC space (cont.)

EXERCISE

A

$$TPR = \frac{63}{63 + 37} = .63$$
$$FPR = \frac{28}{28 + 72} = .28$$

B

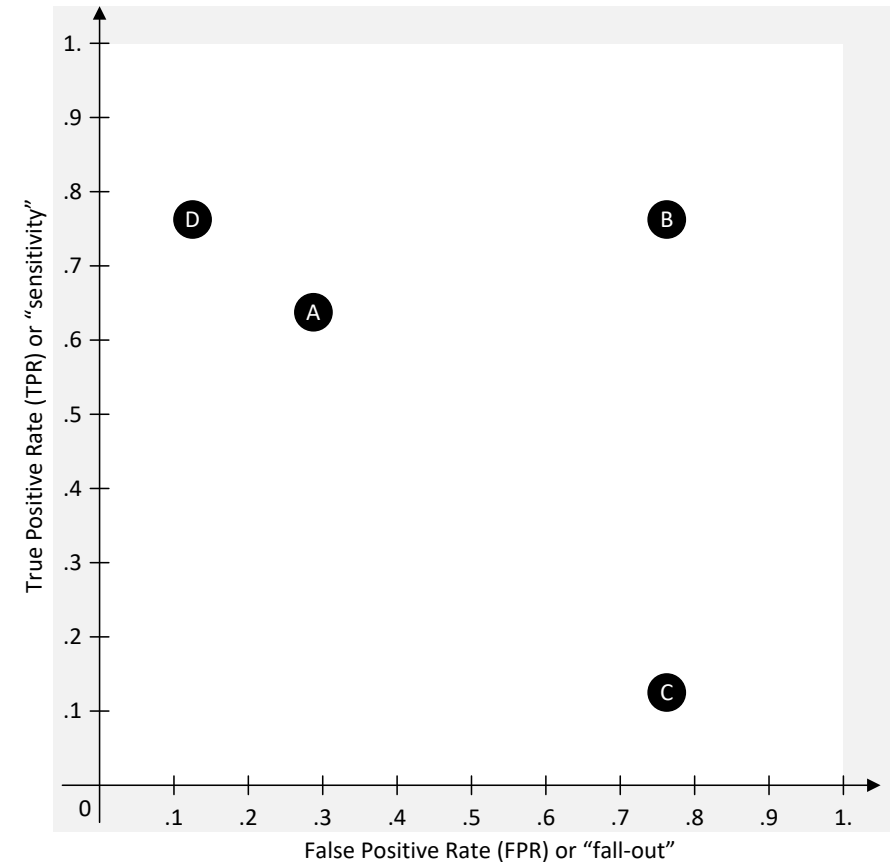
$$TPR = \frac{77}{77 + 23} = .77$$
$$FPR = \frac{77}{77 + 23} = .77$$

C

$$TPR = \frac{24}{24 + 76} = .24$$
$$FPR = \frac{88}{88 + 12} = .88$$

D

$$TPR = \frac{76}{76 + 24} = .76$$
$$FPR = \frac{12}{12 + 88} = .12$$

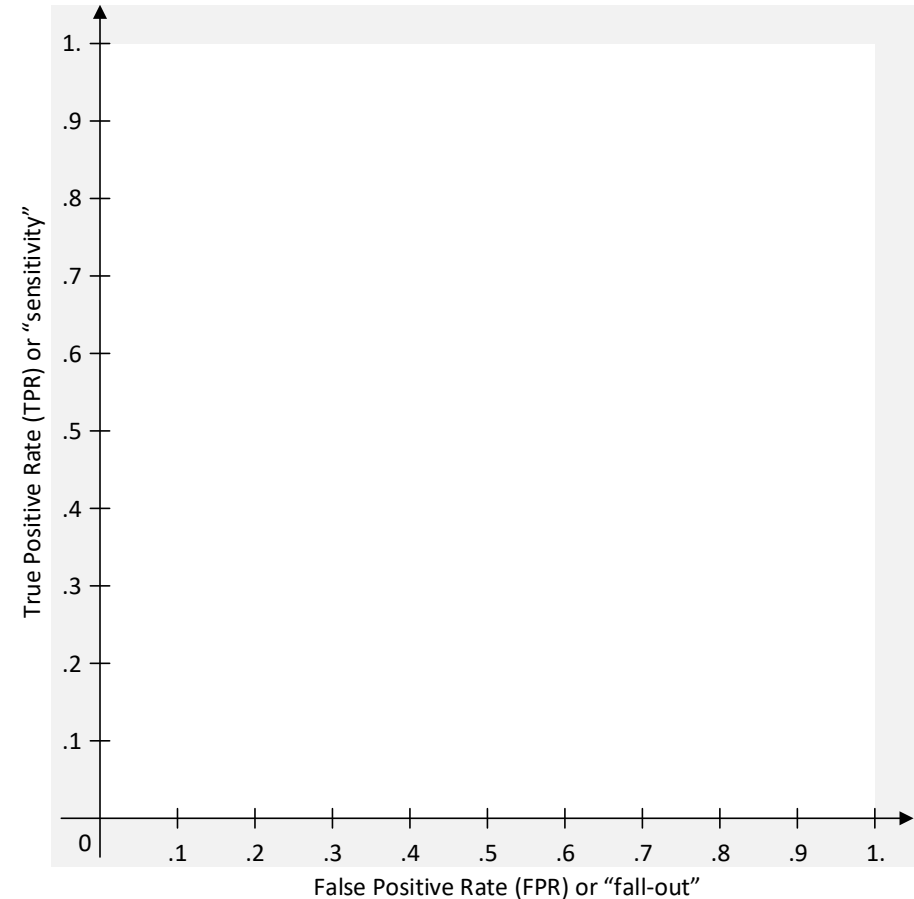


⑥ Build a Model

ROC (receiver operating characteristic or relative operating characteristic) curve

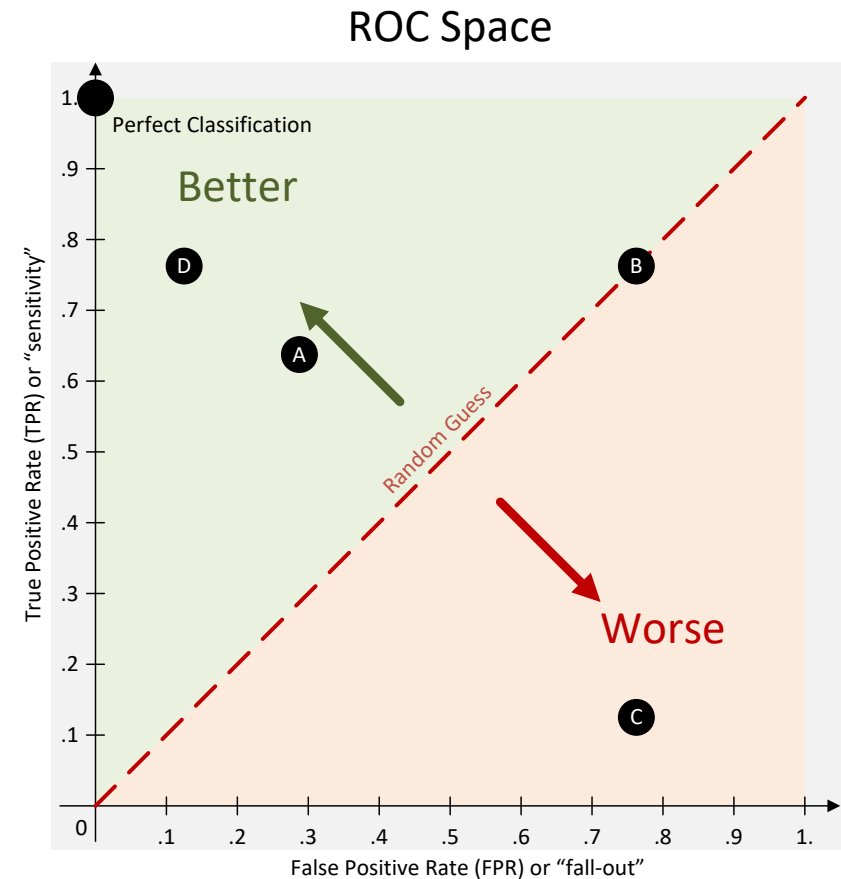
ROC (receiver operating characteristic) curve (a.k.a., relative operating characteristic curve)

- An ROC curve plots the true positive rate (TPR) (or “sensitivity”) against the false positive rate (FPR) (or “fall-out”) at various threshold settings to illustrate the performance of a binary classifier system. The ROC curve is thus the sensitivity as a function of fall-out



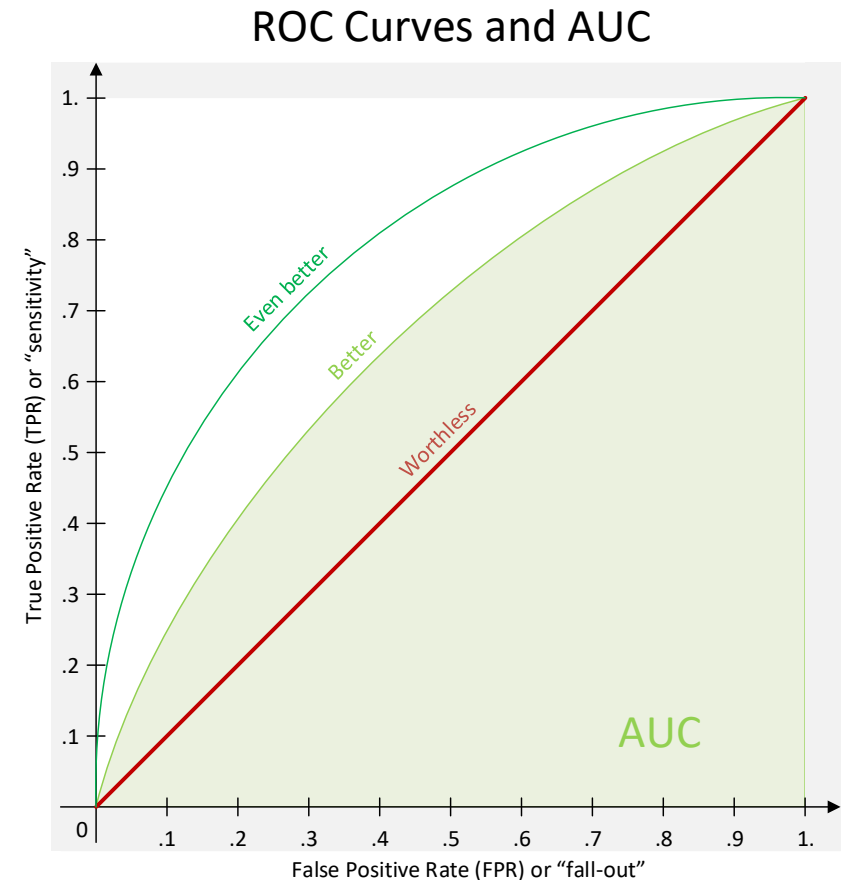
The ROC space demonstrates several things:

- ▶ It shows the tradeoff between sensitivity and fall-out (any increase in sensitivity will be accompanied by an increase in fallout)
 - ▶ The closer the **points** are in the left-hand border and then the top border of the ROC space, the more accurate the classifier is
 - ▶ The closer the **points** come to the 45-degree diagonal of the ROC space, the less accurate the classifier is



The ROC curves demonstrate several things:

- The area under the curve (AUC) is a measure of classifier accuracy
 - The closer the **curve** follows the left-hand border and then the top border of the ROC space, the more accurate the classifier is
 - The closer the **curve** comes to the 45-degree diagonal of the ROC space, the less accurate the classifier is



DS

⑥ Build a Model

Plotting an ROC curve

Plotting an ROC curve

- ▶ ❶ Discard \hat{c} (hypothesized class) and whether it is a true/false positive/negative
- ▶ ❷ Order the trained sample by their decreasing hypothesized probabilities \hat{p} (from more confident to have a '1' down to less confident to have a '1')
- ▶ ❸ Discard the original ranking from the dataset as well as \hat{p}
- ▶ ❹ Start at $(0, 0)$
- ▶ ❺ For each training sample in the sorted order
 - ▶ If $c = 1$, move up by $1/P$
 - ▶ If $c = 0$, move up by $1/N$
- ▶ ❻ If not already at $(1, 1)$, go all the way to the right, then up all the way to $(1, 1)$

Let's plot the ROC for the following trained binary classifier

EXAMPLE

#	\hat{p}	\hat{c}	c	<i>True/False Positive/Negative</i>
1	.44	0	1	FN
2	.29	0	0	TN
3	.98	1	1	TP
4	.69	1	0	FP
5	.07	0	1	FN

❶ Discard \hat{c} (hypothesized class) and whether it is a true/false positive/negative

EXAMPLE

#	\hat{p}	\hat{c}	c	<i>True/False Positive/Negative</i>
1	.44	0	1	FN
2	.29	0	0	TN
3	.98	1	1	TP
4	.69	1	0	FP
5	.07	0	1	FN

② Order the trained sample by their decreasing hypothesized probabilities \hat{p} (from more confident to have a '1' down to less confident to have a '1')

EXAMPLE

#' (ranking by decreasing probabilities)	# (ranking from dataset)	\hat{p}	c
1	3	.98	1
2	4	.69	0
3	1	.44	1
4	2	.29	0
5	5	.07	1

③ Discard the original ranking from the dataset as well as \hat{p}

EXAMPLE

#' (ranking by decreasing probabilities)	# (ranking from dataset)	\hat{p}	c
1	3	.98	1
2	4	.69	0
3	1	.44	1
4	2	.29	0
5	5	.07	1

Let's plot the ROC/AUC for the following trained binary classifier (cont.)

EXAMPLE

#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1

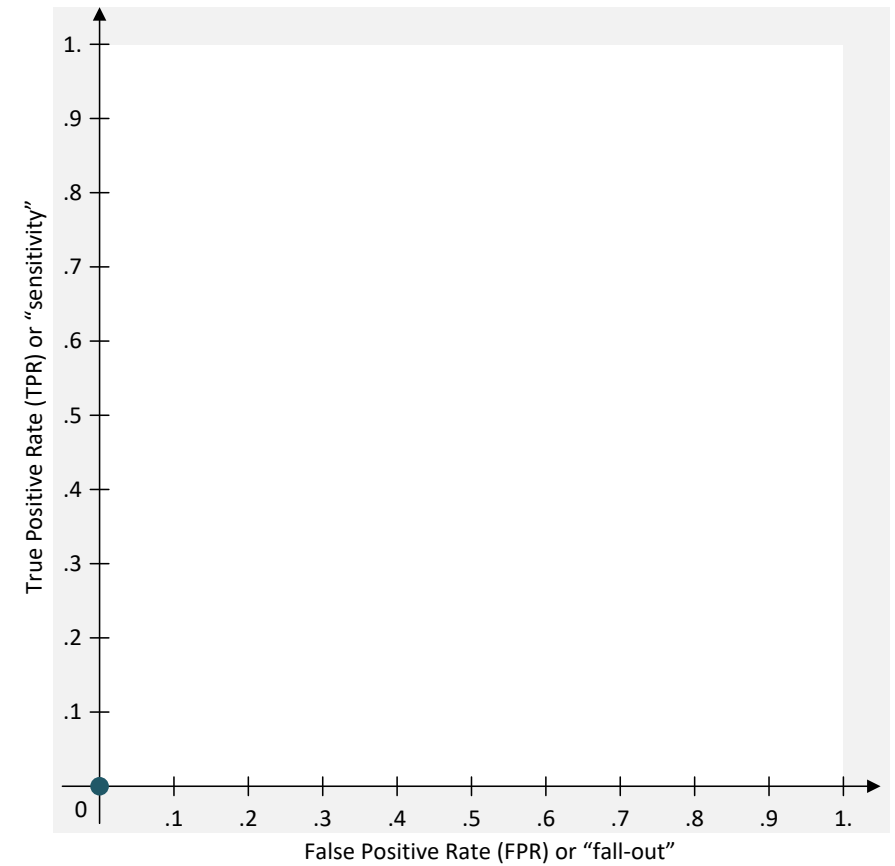
▸ $P = 3 \rightarrow 1/P = 1/3$

▸ $N = 2 \rightarrow 1/N = 1/2$

④ Start at (0, 0)

EXAMPLE

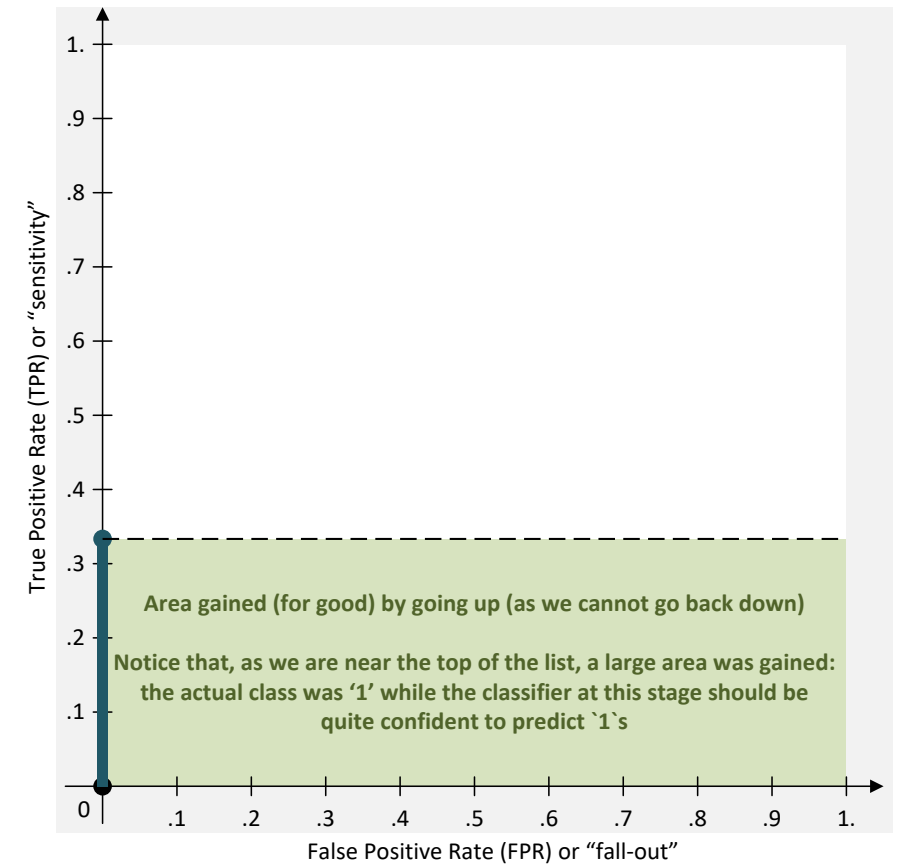
#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1



⑤ Because $c = 1$, move up by $1/P = 1/3$

EXAMPLE

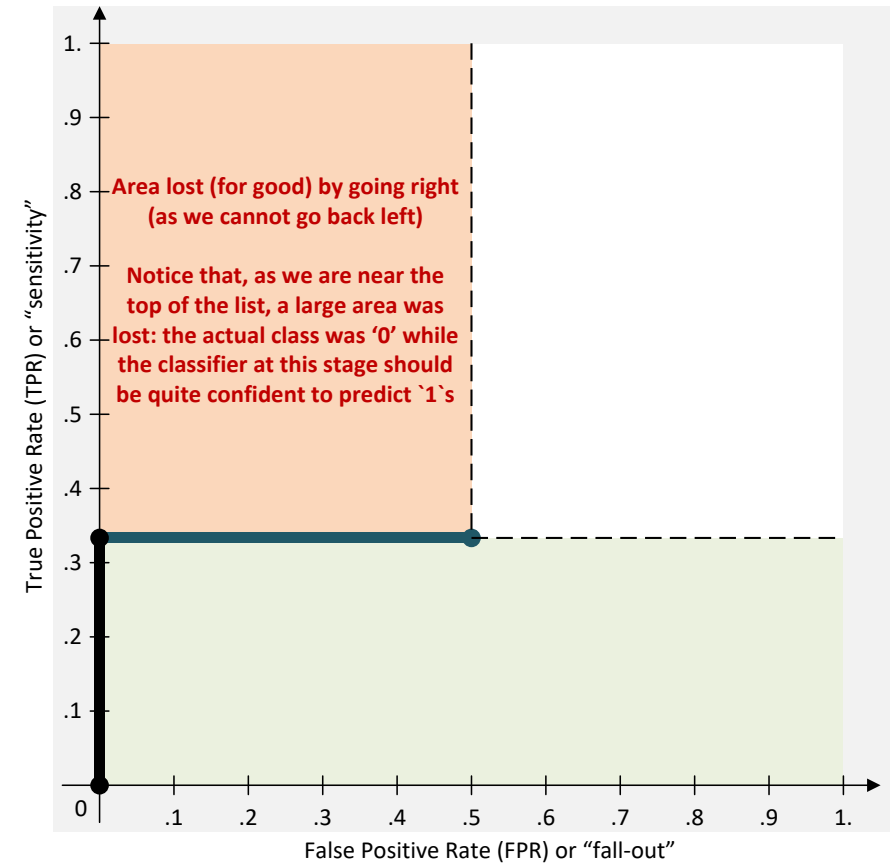
#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1



⑤ Because $c = 0$, move right by $1/N = 1/2$

EXAMPLE

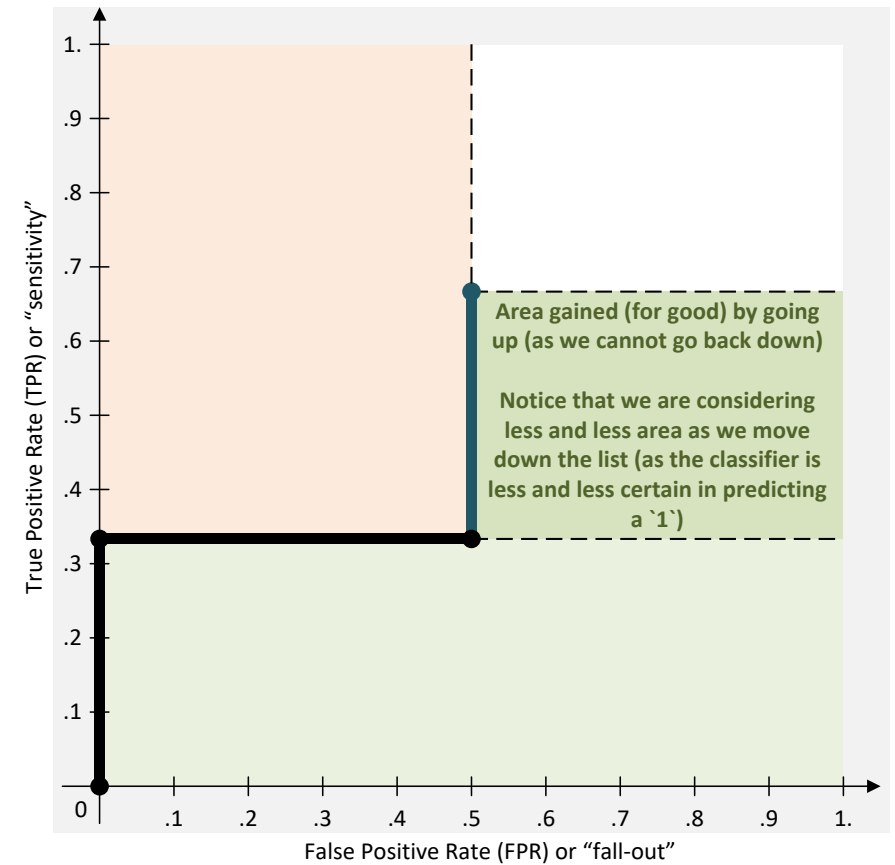
#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1



⑤ Because $c = 1$, move up by $1/P = 1/3$

EXAMPLE

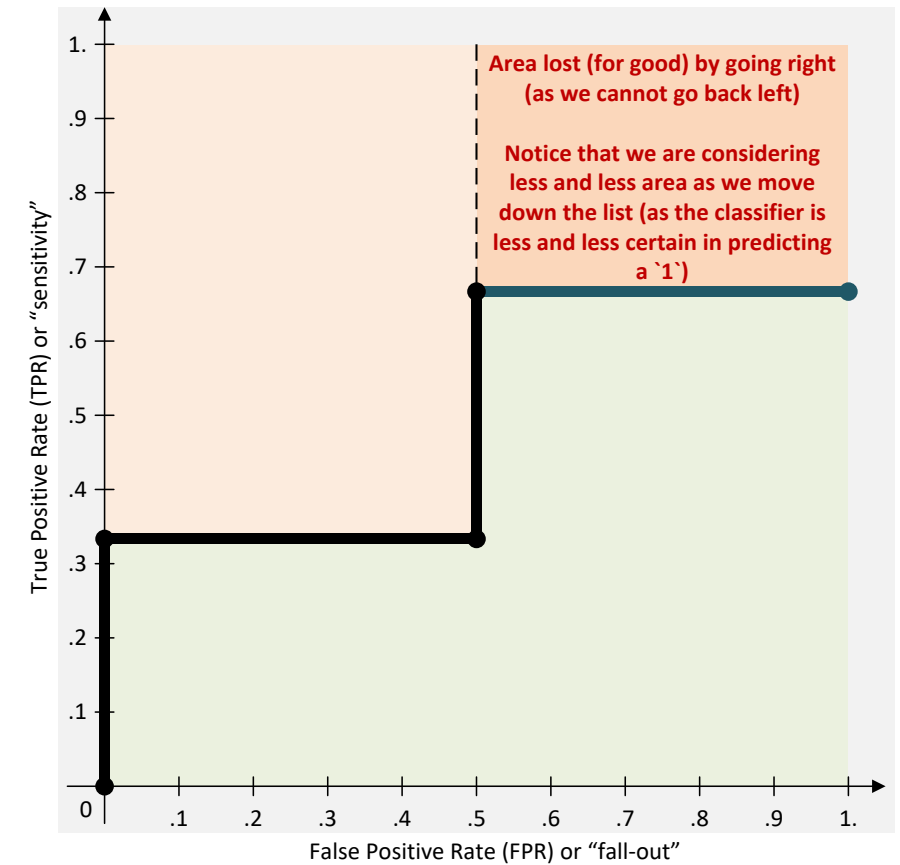
#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1



⑤ Because $c = 0$, move left by $1/N = 1/2$

EXAMPLE

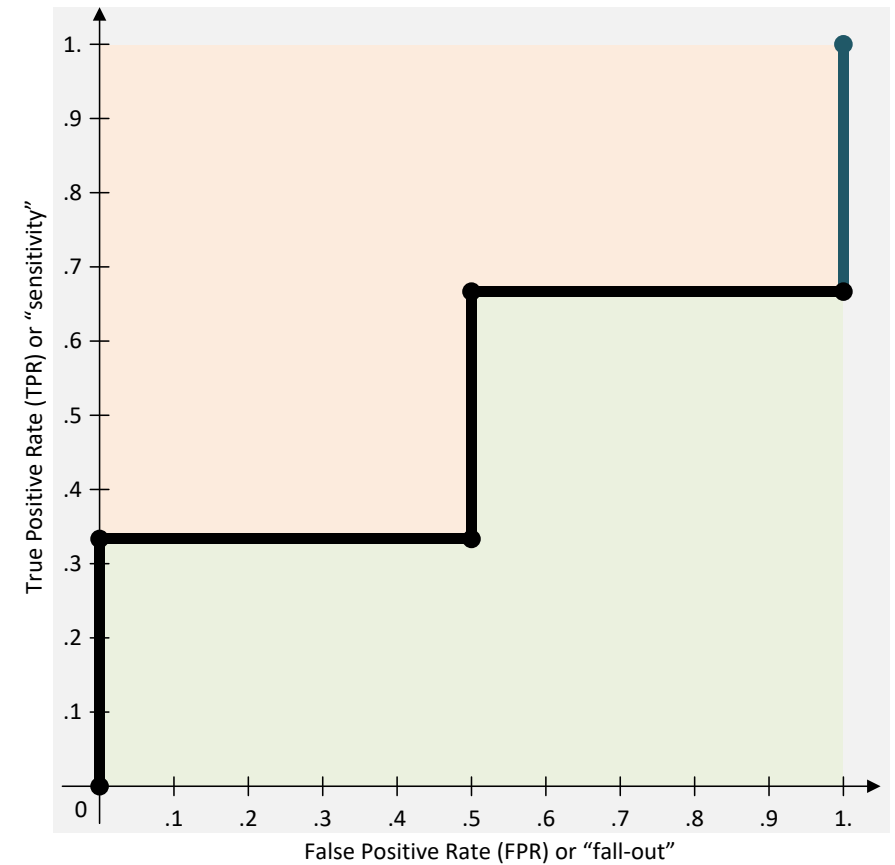
#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1



⑤ Because $c = 1$, move up by $1/P = 1/3$

EXAMPLE

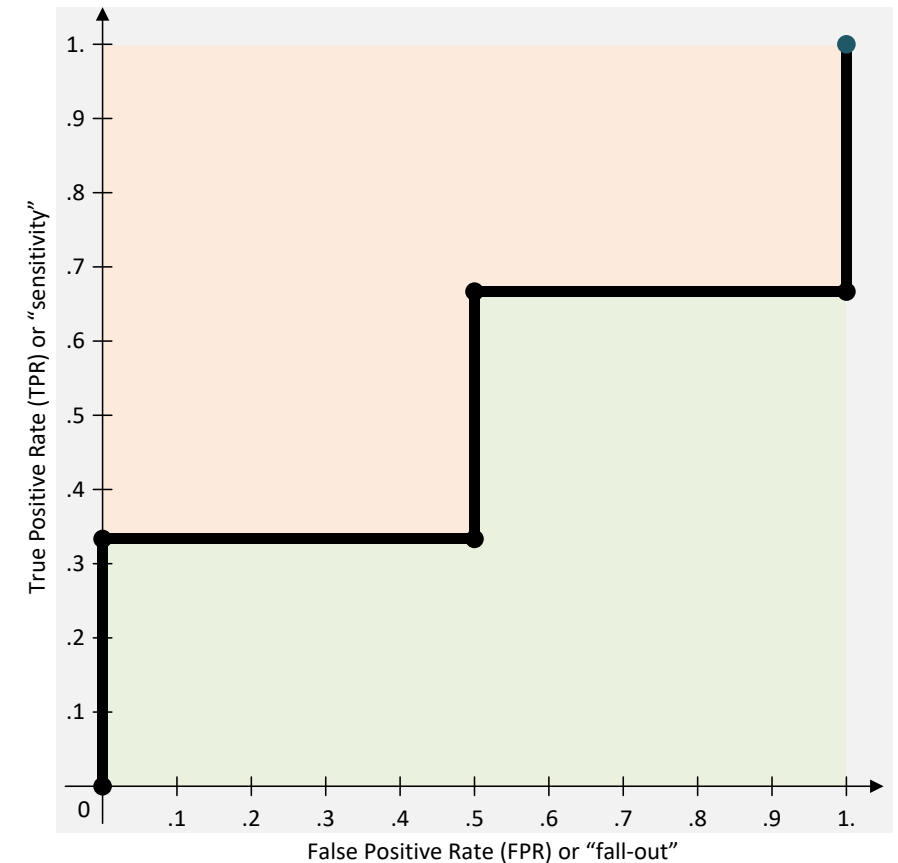
#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1



⑥ If not already at (1, 1), go all the way to the right, then up all the way to (1, 1)

EXAMPLE

#' (ranking by decreasing probabilities)	c
1	1
2	0
3	1
4	0
5	1



Plotting an ROC curve (cont.)

- Notes

- We don't rely on a threshold (e.g., .5) for plotting ROC curves. Indeed, moving up or right is independent of \hat{p} (we discarded it in step ③) and only relies on a decreasing ranking of \hat{p} and then c
- As a matter of fact, you can use ROC curves to select the best threshold but we won't address it here

DS

⑥ Build a Model

Codealong – Part B

ROC/AUC

DS

7 Present the Results

Communicating Results

We built a model! Now what?

- We've built our model, but there is still a gap between our iPython notebook with its plots and figures and a slideshow needed to present our 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? (cont.)

- 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? (cont.)

- Who is your audience? Are they technical? What are their concerns?
 - 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? (cont.)

- Today, we'll focus on communicating results for “simpler” problems, but this applies to any type of model you may work with

DS

7 Present the Results

Showing our Work

Showing our 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

To convey important information to your audience, make sure your charts are simplified, easily interpretable, and 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 against a target
- If your model used a data transformation (like natural log), just visualize the original data
- Remove 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

When building visuals for another audience, ask yourself who, what, and how

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?

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
- E.g., in logistic regression, plotting the probability of a class against a variable can help explain the range of effect of the model

Visualizing Performance against Baseline

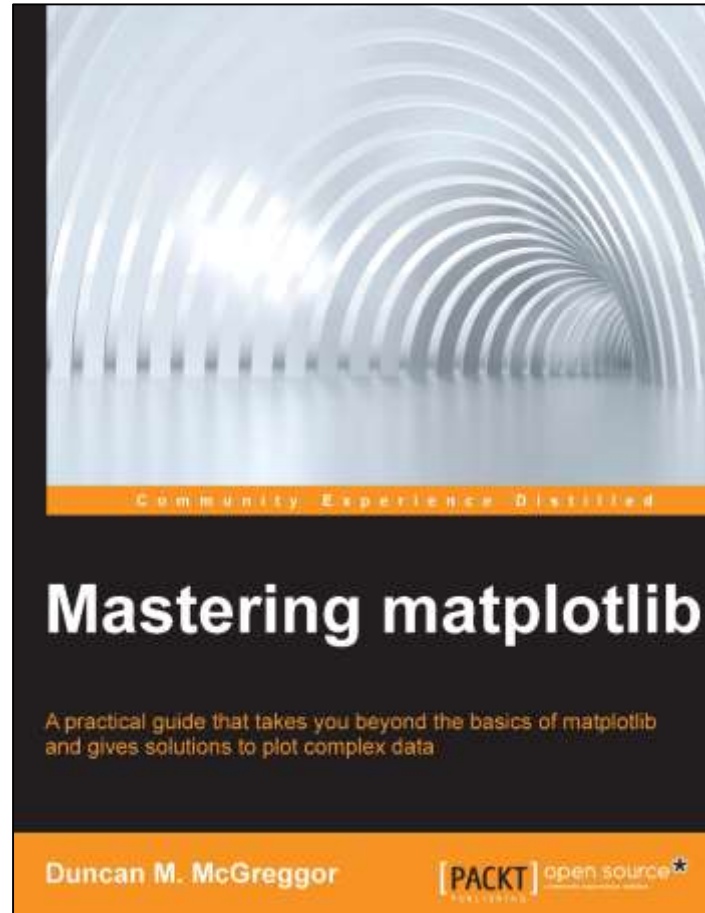
- Another approach of visualization is the effect of your model against a baseline, or – even better – against previous models
- Plots like this will also be useful when talking to your peers – other data scientists or analysts who are familiar with your project and interested in the progress you've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

DS

7 Present the Results

Codealong – Part C
Prettying up Graphs

A good resource to learn more about *matplotlib*
(optional; not required for the course)





DS

Review

Review

You should now be able to:

- Evaluate a model using advanced metrics such as confusion matrix and ROC/AUC curves
- Explain the trade-offs between false positives vs. false negatives
- Describe the difference between visualization for presentations vs. exploratory data analysis
- Identify the components of a concise, convincing report and how they relate to specific audiences/stakeholders

Next Class

Decision Trees and Random Forests

Learning Objectives

After the next lesson, you should be able to:

- Understand and build decision tree models for classification and regression
- Understand the differences between linear and non-linear models
- Understand and build random forest models for classification and regression
- Know how to extract the most important predictors in a random forest model



DS

Exit Ticket

Don't forget to fill out your exit ticket [here](#)

Slides © 2016 Ivan Corneillet Where Applicable
Do Not Reproduce Without Permission