Regression vs Classification

UW
DATA SCIENCE
CLUB.

Presented by Neysa Patel

Workshop Overview

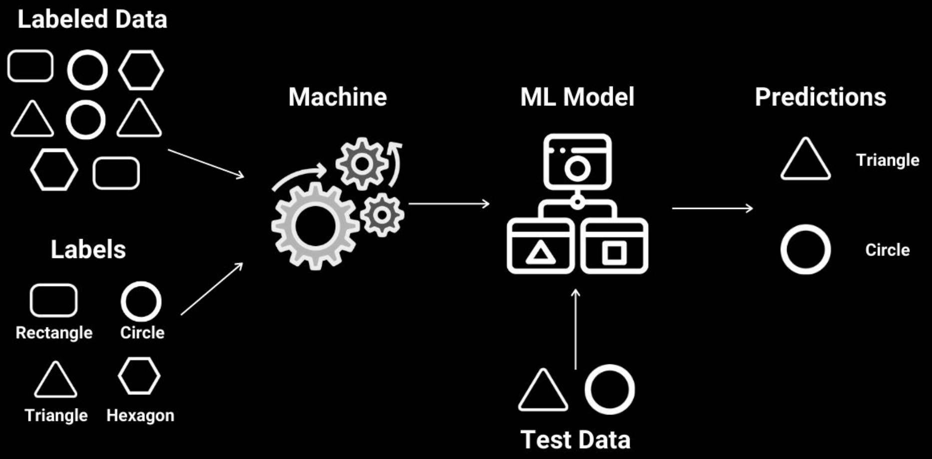
Goals:

- Gain an understanding of what regression and classification are, and when to use which one
- Learn about some examples of regression and classification in practice
- Techniques to use regression and classification to extract meaningful insights from data

Supervised Learning

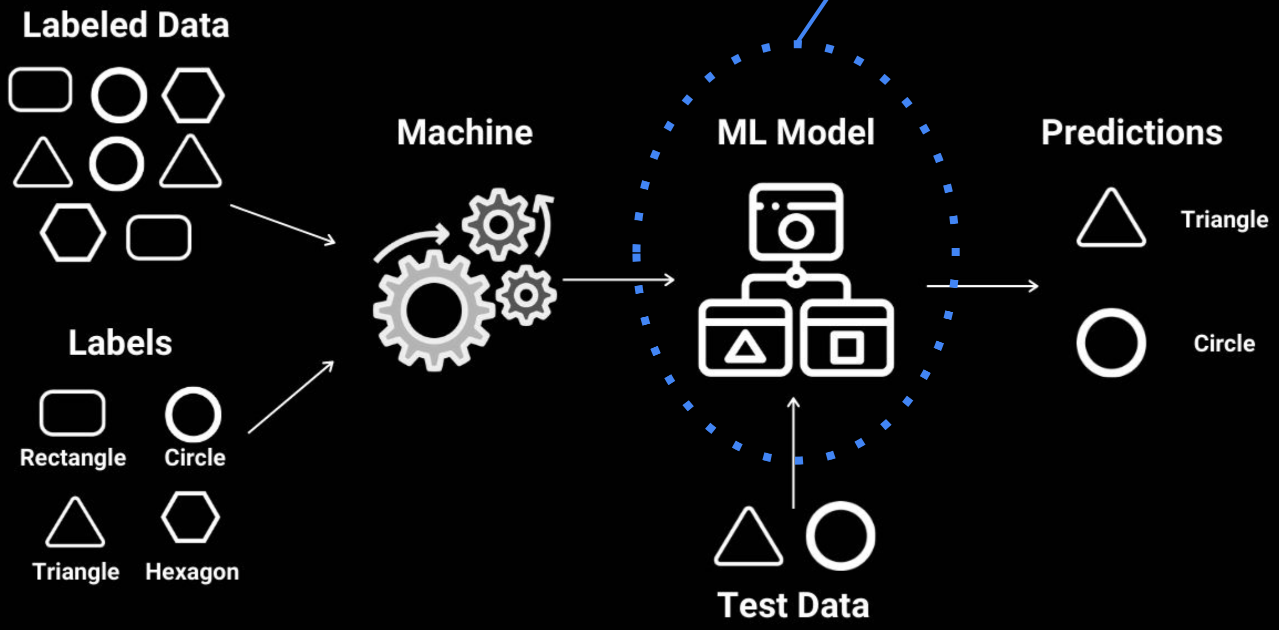
- Learn the relationship between input and output through labelled training data

- Given a set of pre-classified training examples, classify a new instance



Predictions?

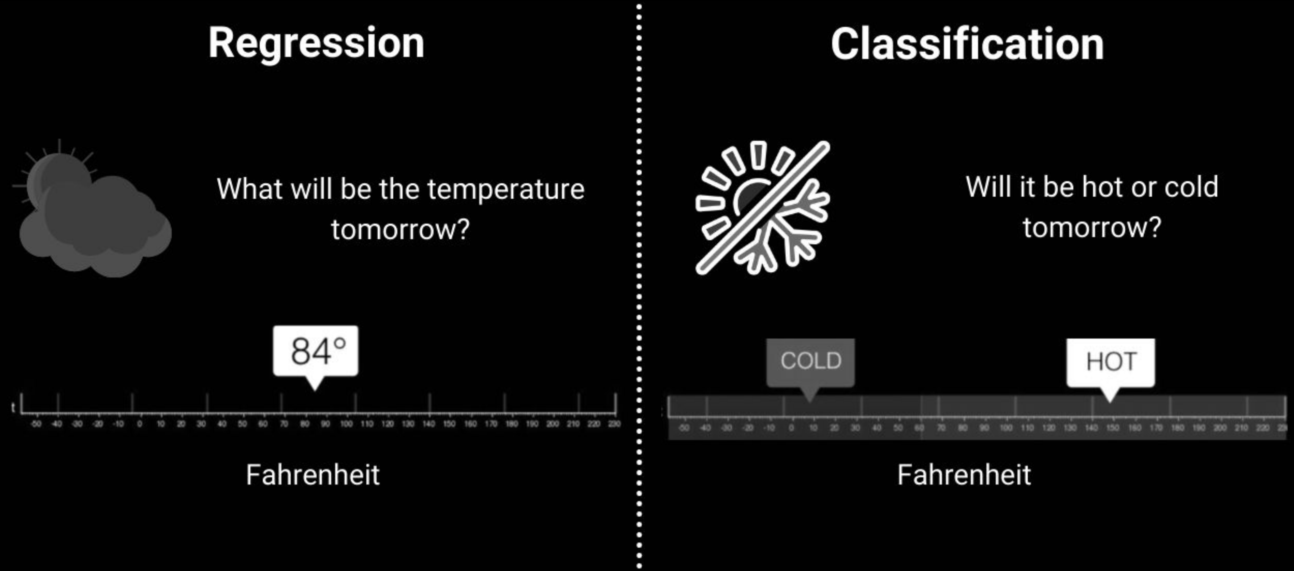
How does the model know what to predict the shape as?



UW DSC.

Predictive Analysis

- Regression and Classification!



Regression

UW DSC.

- Used for continuous (numerical) data
- Predict “the value” of something given its past “values”
  - What will the MSFT stock price be when the market closes tomorrow?
  - What will the temperature be at 3 PM next Friday?
  - How much could I sell my house for, 10 years from now?

Regression

UW DSC.

- Associated with each “value” are a set of “features,” which maybe you can use to predict your “value”

Distance from Toronto (km)	Age (Years)	Square Footage	Bedrooms	Number of Purple Walls in House	Price ($)
249	10	2100	3	2	824 920
16	5	1700	2	0	1 439 014

- Do all these features contribute equally to determine the house price?

Regression

UW
DSC.
–

- Do all these features contribute equally to determine the house price?
  - No! Every feature has a certain “weight”
  - Maybe the square footage of the house matters the most and the number of purple walls in the house matters the least

Distance from Toronto (km)	Age (Years)	Square Footage	Number of Bedrooms	Number of Purple Walls in House	Price ($)
249	10	2100	3	2	824 920
16	5	1700	2	0	1 439 014

Regression

UW
DSC.
–

- Determine which features, in which combination, can predict the value!

Distance from Toronto (km)	Age (Years)	Square Footage	Number of Bedrooms	Number of Purple Walls in House	Price ($)
249	10	2100	3	2	824 920
16	5	1700	2	0	1 439 014

Linear Regression

UW DSC.

- Dependency between variables is linear in terms of inputs
  - In our example, we have 5 variables:
    - X_1 = Distance from Toronto
    - X_2 = Age
    - X_3 = Square footage
    - X_4 = Number of bedrooms
    - X_5 = Number of purple walls in the house

Linear Regression

UW DSC.

- In our example, each variable has a weight associated with it, for how important of a factor it is in determining the house price:
  - β_1 = Weight of X_1 (Distance from Toronto)
  - β_2 = Weight of X_2 (Age)
  - β_3 = Weight of X_3 (Square footage)
  - β_4 = Weight of X_4 (Number of bedrooms)
  - β_5 = Weight of X_5 (Number of purple walls in the house)

Linear Regression

UW
DSC.

- Tying this all together, we can represent the cost of the house as:

    - $y = (\beta_1 \cdot X_1) + (\beta_2 \cdot X_2) + (\beta_3 \cdot X_3) + (\beta_4 \cdot X_4) + (\beta_5 \cdot X_5) + \varepsilon$

- Intuition: $y = mx + b$
    - Except, there is more than 1 independent variable (X's)

UW
DSC.
–

Linear Regression

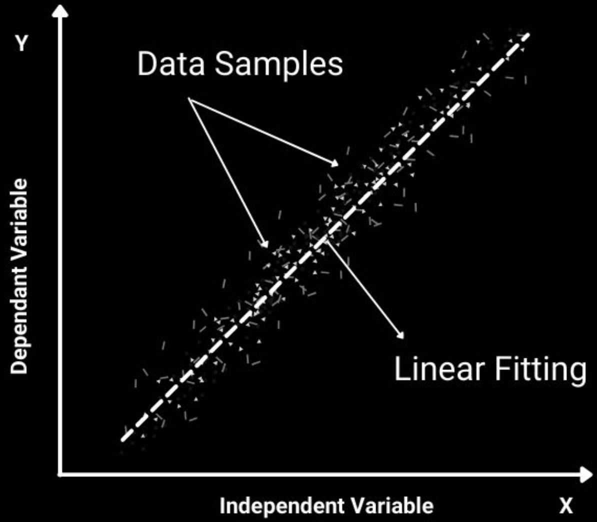
- Tying this all together, we can represent the cost of the house as:
  - $y = (\beta_1 \cdot X_1) + (\beta_2 \cdot X_2) + (\beta_3 \cdot X_3) + (\beta_4 \cdot X_4) + (\beta_5 \cdot X_5) + \varepsilon$

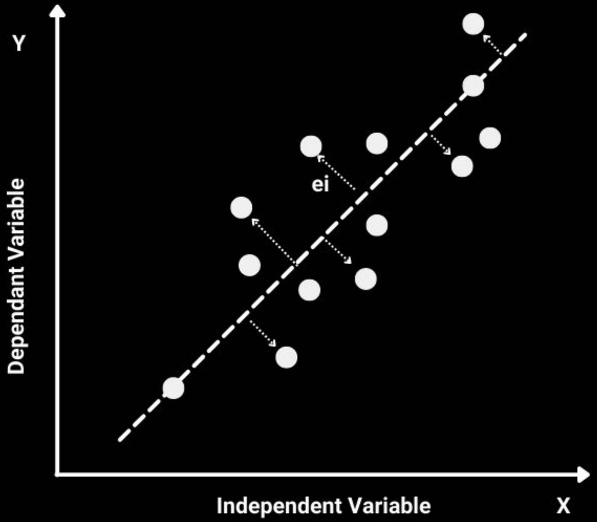
- ε is the error due to fitting imperfection, since we can't assume that all data samples will follow the expected function *perfectly*

Linear Regression

UW
DSC.

- Learn the linear relationship between one (or more) input features (X) and the single output variable (Y) based on historical data





Loss Function in Linear Regression

UW DSC.

- Suppose the actual value for input A is Y, and our linear regression model predicted Y’ for the same input A

- The error for A is:

  ei = |y - y’| = (y - y’)^2

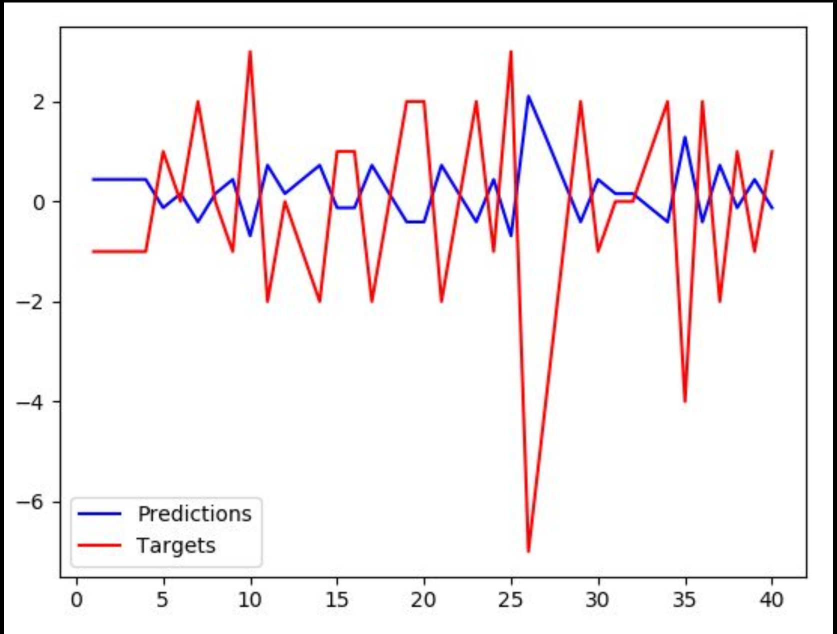
- This is for one sample A. So the cumulative error for all the samples in the dataset is the *sum of square residuals*:

$$f(x^{(j)}) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i^{(j)}$$

$$SSR = \sum_{j=1}^{m} (y^{(j)} - f(x^{(j)}))^2$$

Regression Model Evaluation

- How good is our model?



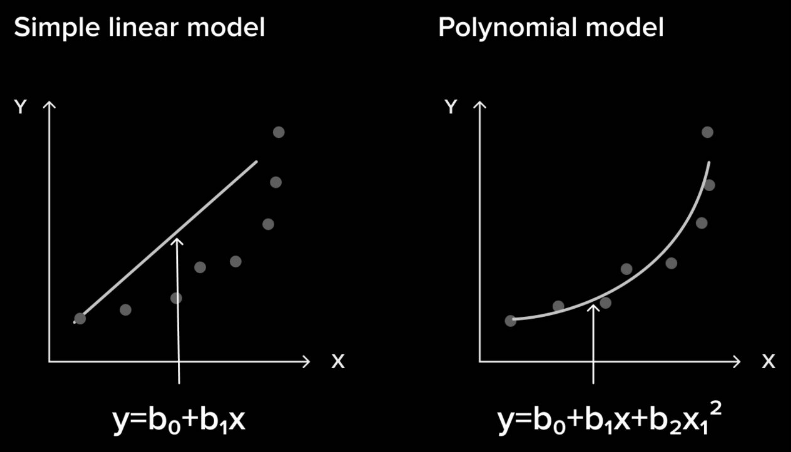
Regression Model Evaluation

UW DSC.

- Mean Squared Error
  - Measures the average squared difference between actual and predicted values
- R^2
  - Quantifies how well the regression model fits the data, with a higher R^2 indicating a better fit

Regression

- Sometimes the relationship between the independent and dependent variables is too complicated to be described by a linear relationship
  - Use a higher order (polynomial) function in this case



Regression

UW DSC.

- What are some of the “features” (independent variables) for our examples earlier?
  - What will the MSFT stock price be when the market closes tomorrow?
  - What will the temperature be at 3 PM next Friday?

Classification

UW DSC.

- Used for discrete (categorical) data
- Predict “the class” of something given its past categories (it falls into predefined classes or categories)
    - Is this email spam or not spam, based on its content?
    - I have a fruit. Is it an apple, a banana, or a cherry?
    - Will it be hot or cold tomorrow?
    - Can I sell my house for over $1 million next year?

Classification

UW DSC.

- Associated with each “value” are a set of “features,” which maybe you can use to predict your “value”

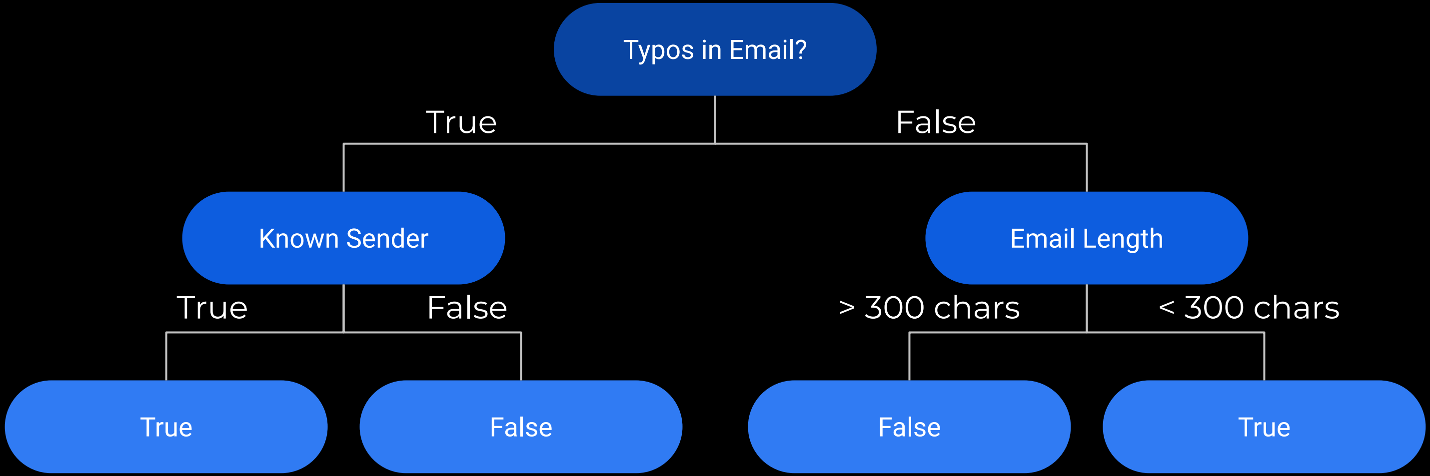
Typos in the Email?	**Email Length (chars)**	**Known Sender**	**.edu domain?**	**Spam?**
True	49	False	True	True
False	272	True	False	False

- Do all these features contribute equally to determine if the email is spam or not?

Decision Trees

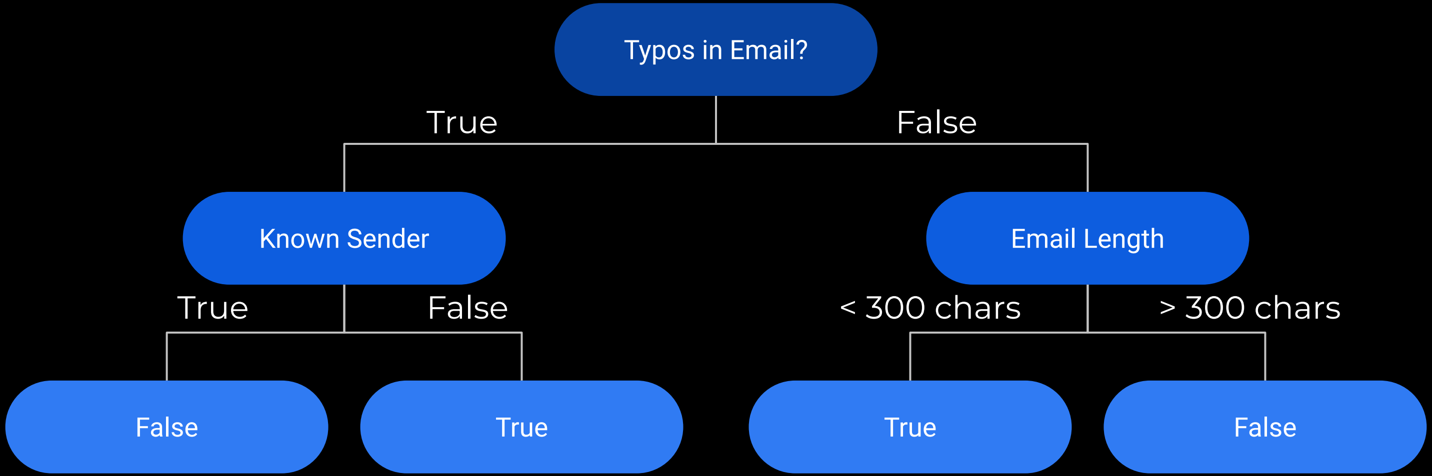
UW
DSC.

- Follow the tree until you reach a decision (classification)!



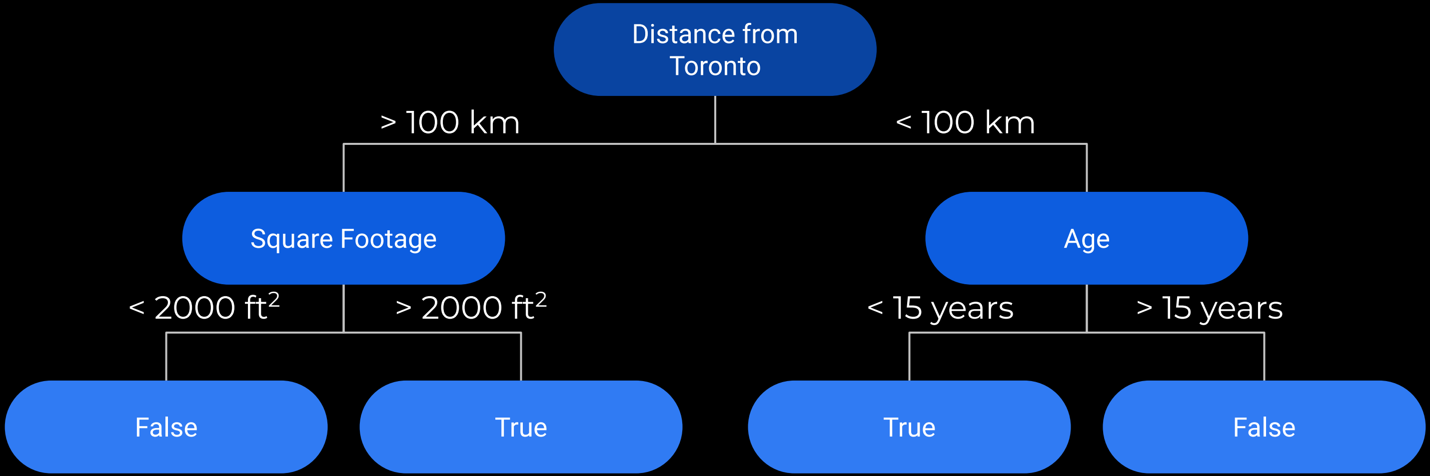
Decision Trees

UW DSC.



Typos in the Email?	Email Length (chars)	Known Sender	.edu domain?	Spam?
True	103	False	True	??

Decision Trees

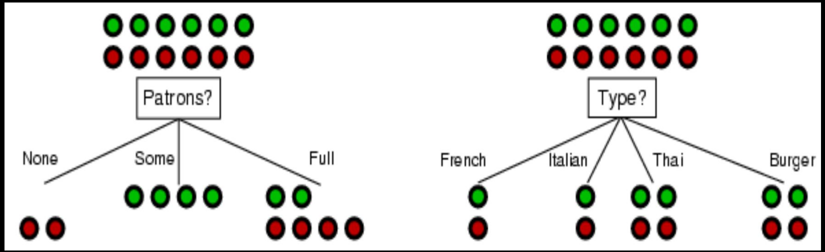


Distance from Toronto (km)	Age (Years)	Square Footage	Bedrooms	Number of Purple Walls in House	Price Over $1 Million?
55	20	2600	4	13	??

Decision Trees: Choosing an Attribute

- There are different ways of selecting attributes, but generally a “good attribute” splits the training examples appropriately

Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*Wait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T



Source: CS 486 Lecture Notes

ROC + F-1 Score

- How “good” is our classifier?

		Actual: Positive	Actual: Negative
Prediction	Positive	TP	FP
	Negative	FN	TN

UW
DSC.

ROC + F-1 Score

$$Sensitivity = \frac{TP}{TP+FN}$$

$$Specificity = \frac{TN}{TN+FP}$$

$$Precision = \frac{TP}{TP+FP}$$

$$F - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

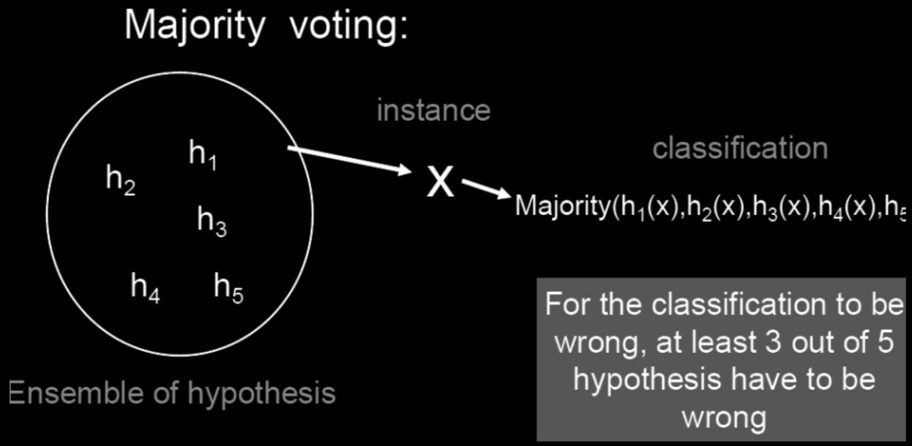
		Actual	
		Positive	Negative
Prediction	Positive	TP	FP
	Negative	FN	TN

Ensembles

- Intuition:
  - Individuals may make mistakes, but the majority may be less likely to make a mistake
  - Individuals have partial information but committees can pool their expertise

Bagging

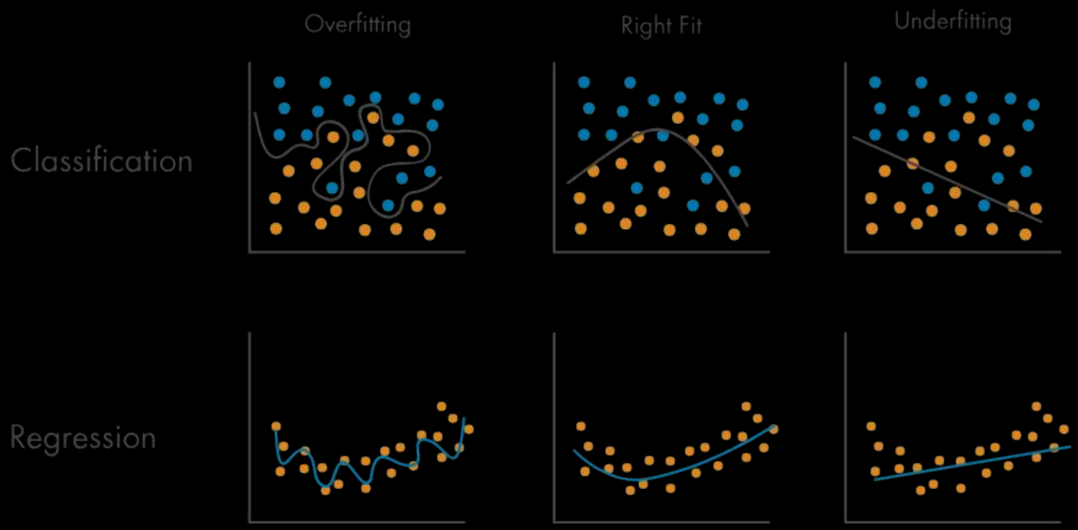
- Random Forests
  - Many unique decision trees classify instance X
  - Classification = what most trees classified X as



Source: CS 486 Lecture Notes

Overfitting

- Finding patterns in the data where there is no actual pattern
- Bias!



Quiz!

UW
DSC.