L4: ML evaluation

Prof. Xun Jiao

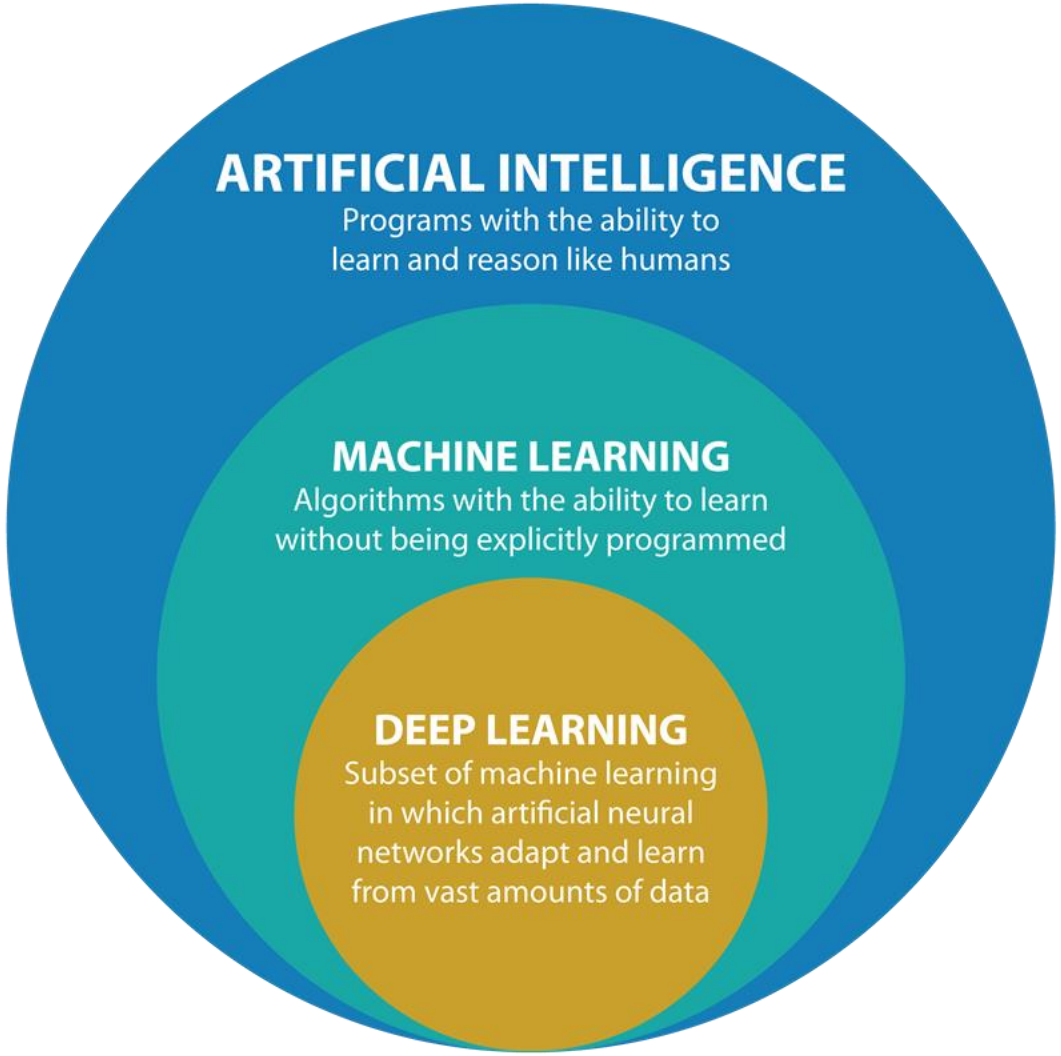
Before class

- Slides uploaded to BB
- HWs this week
  - Will be posted on Wed
  - ML evaluation: A ML model without a proper evaluation is useless
  - Due by next Wed
  - Detailed instruction will be on BB.

Review

- AI VS. ML VS. DL
  - What are two different types of ML?
    - What are two different types of supervised learning?

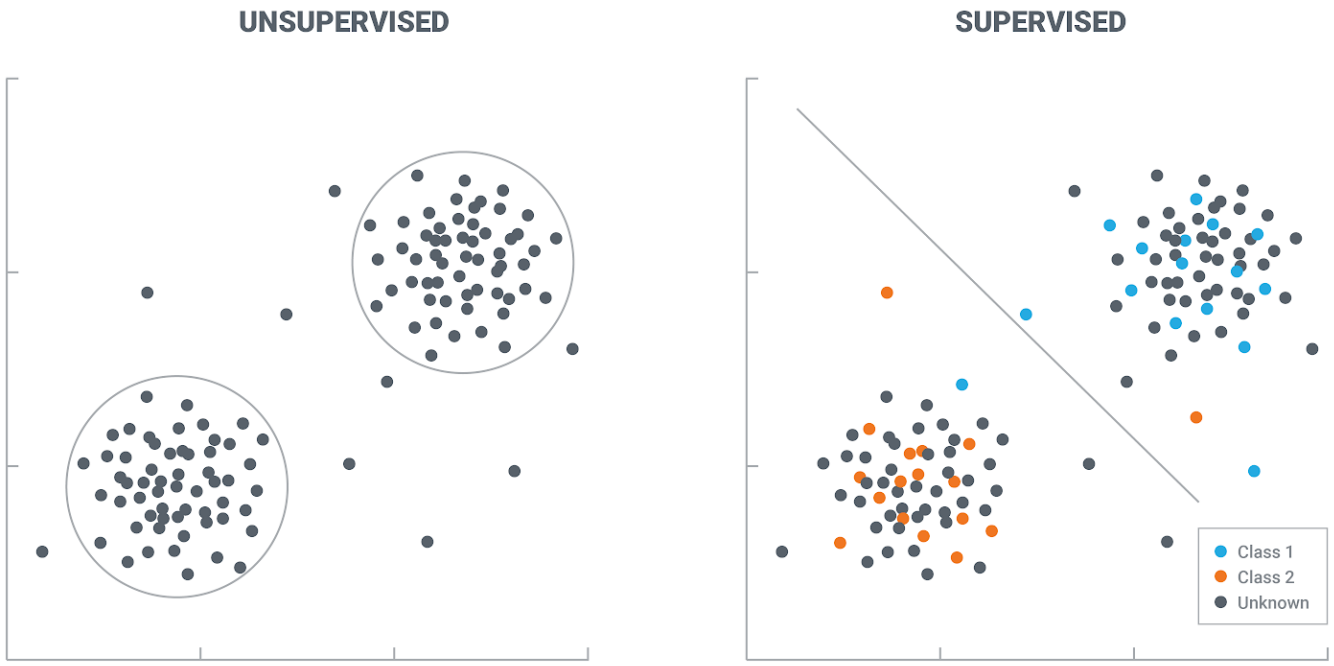
AI VS. Machine Learning



- ***Artificial intelligence*** is a wide field, which aims at making machines intelligent. AI has a set of tools through which it enables a machine to mimic human intelligence.
  - Natural language processing
  - Robotics
  - Machine Learning
  - Self-driving cars
  - https://en.wikipedia.org/wiki/Artificial_intelligence#Reasoning,_problem_solving

- One of the tools AI have is ***machine learning***, that gives the machines to learn without being told explicitly what to do.

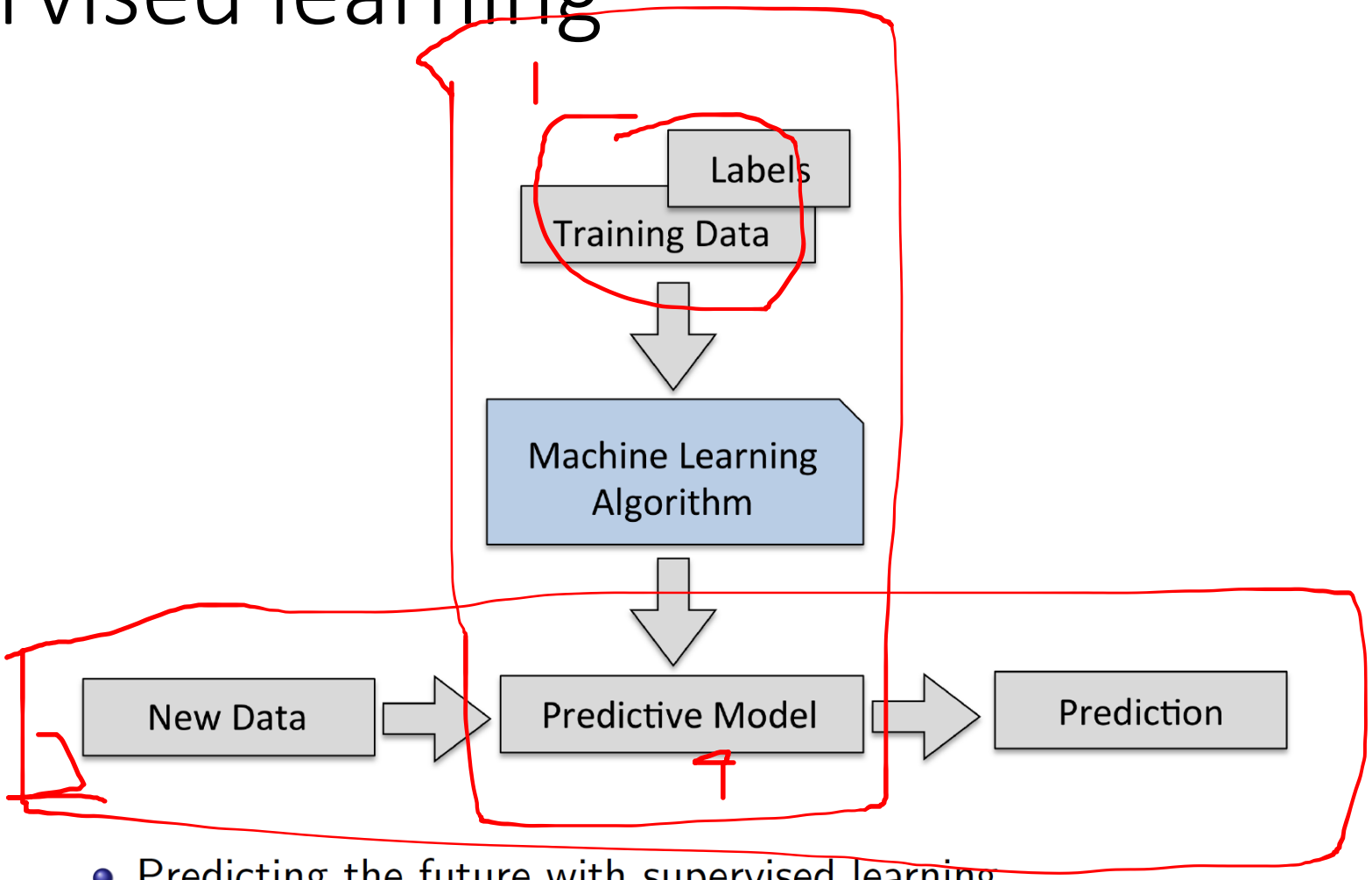
- Machine learning again has various tools in its pocket, one of them being neural networks. ***Neural networks*** try to mimic the activity of a human brain. ***Deep learning*** is the use of more sophisticated neural networks, with more non-linear layers, convolutional layers etcetera.



Two classes of ML: supervised VS. unsupervised

- Supervised learning:
  - Goal: perform task as good as humans
  - Experience: training data provided by human
  - Performance: accuracy/error on the task
- Unsupervised learning:
  - Goal: find some structure in data
  - No experience
  - No (explicit) performance metric

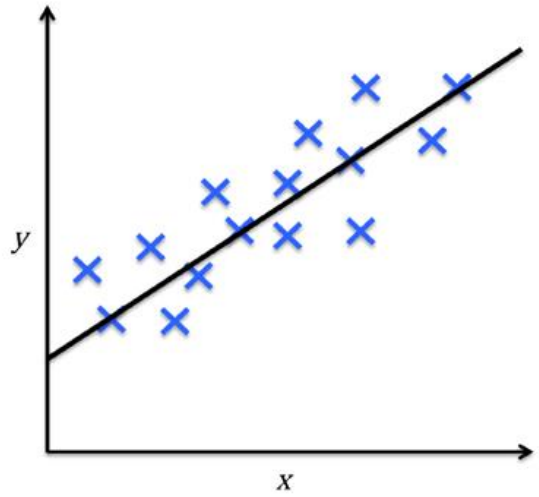
Supervised learning



- Predicting the future with supervised learning

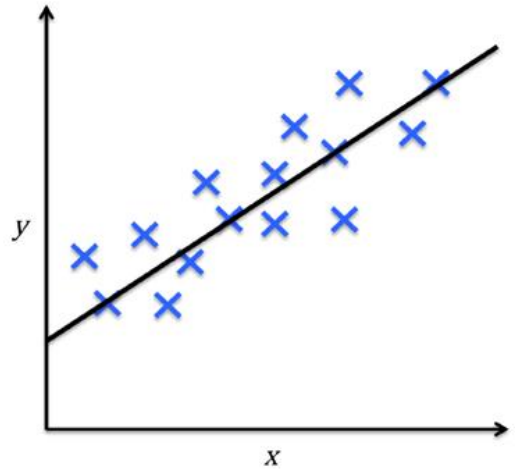
Supervised learning is again classified as: regression VS. classification

- Regression predictive modeling
  - Regression is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y).
    - A continuous output variable is a real-value, such as an integer or floating point value. These are often quantities, such as amounts and sizes.
      - For example, a house may be predicted to sell for a specific dollar value, e.g., $300k, $350k, $390k, etc.
      - Student grades.



Supervised learning is again classified as: regression VS. classification

- Regression predictive modeling
  - A regression problem requires the prediction of a quantity.
  - A prediction cannot be perfect
    - There are always errors
    - How to evaluate performance of ML?
  - Evaluation metric: deviation from golden results
    - Mean absolute error, mean relative error, MSE (mean squared error), RMSE (root mean squared error)



Mean absolute error (MAE)

- MAE = Average of All absolute errors

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y'_i - y_i|$$

No.	Model	Luxury?	Year	MPG	Horsepower	Price	Model 1	Model 2
1	Acura MDX	Yes	2017	20	290	$50,000	46,000	45,000
2	Honda Accord	No	2017	25	190	$25,000	27,000	26,000
3	Honda Civic	No	2012	23	160	$10,000	9,900	12,000
4	Honda Civic	No	2016	24	170	$18,000	19,000	19,000
5	Nissan Altima	No	2016	30	180	$25,000	26,000	27,000
6	Acura MDX	Yes	2015	18	280	$38,000	40,000	35,000
7	Lexus RX350	Yes	2015	21	270	$40,000	43,000	36,000
8	Toyota Prius	No	2014	45	120	$28,000	24,000	29,000
9	Toyota Prius	No	2013	40	120	$24,000	23,000	27,000

Which model is better?
What is the best possible MAE?

Mean absolute percentage error (MAPE)

- MAPE focuses more on the "relative" rather than "absolute"

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y'_i - y_i}{y_i}\right|$$

- Anything wrong with the equation?

No.	Model	Luxury?	Year	MPG	Horsepower	Price	Model 1	Model 2
1	Acura MDX	Yes	2017	20	290	$50,000	46,000	45,000
2	Honda Accord	No	2017	25	190	$25,000	27,000	26,000
3	Honda Civic	No	2012	23	160	$10,000	9,900	12,000
4	Honda Civic	No	2016	24	170	$18,000	19,000	19,000
5	Nissan Altima	No	2016	30	180	$25,000	26,000	27,000
6	Acura MDX	Yes	2015	18	280	$38,000	40,000	35,000
7	Lexus RX350	Yes	2015	21	270	$40,000	43,000	36,000
8	Toyota Prius	No	2014	45	120	$28,000	24,000	29,000
9	Toyota Prius	No	2013	40	120	$24,000	23,000	27,000

Which model is better?
What is the best possible MAPE?

Mean squared error

- measures the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

No.	Model	Luxury?	Year	MPG	Horsepower	Price	Model 1	Model 2
1	Acura MDX	Yes	2017	20	290	$50,000	46,000	45,000
2	Honda Accord	No	2017	25	190	$25,000	27,000	26,000
3	Honda Civic	No	2012	23	160	$10,000	9,900	12,000
4	Honda Civic	No	2016	24	170	$18,000	19,000	19,000
5	Nissan Altima	No	2016	30	180	$25,000	26,000	27,000
6	Acura MDX	Yes	2015	18	280	$38,000	40,000	35,000
7	Lexus RX350	Yes	2015	21	270	$40,000	43,000	36,000
8	Toyota Prius	No	2014	45	120	$28,000	24,000	29,000
9	Toyota Prius	No	2013	40	120	$24,000	23,000	27,000

Which model is better?
What is the best possible MSE?

Root mean squared error

- Root of MSE:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

No.	Model	Luxury?	Year	MPG	Horsepower	Price	Model 1	Model 2
1	Acura MDX	Yes	2017	20	290	$50,000	46,000	45,000
2	Honda Accord	No	2017	25	190	$25,000	27,000	26,000
3	Honda Civic	No	2012	23	160	$10,000	9,900	12,000
4	Honda Civic	No	2016	24	170	$18,000	19,000	19,000
5	Nissan Altima	No	2016	30	180	$25,000	26,000	27,000
6	Acura MDX	Yes	2015	18	280	$38,000	40,000	35,000
7	Lexus RX350	Yes	2015	21	270	$40,000	43,000	36,000
8	Toyota Prius	No	2014	45	120	$28,000	24,000	29,000
9	Toyota Prius	No	2013	40	120	$24,000	23,000	27,000

Which model is better?
What is the best possible MAPE?

Use python code to calculate metrics

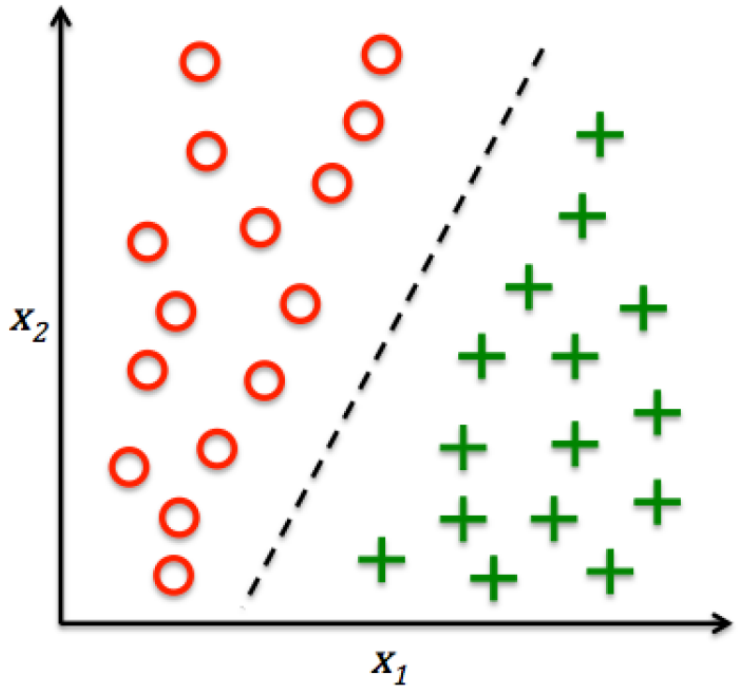
- Let us do some Python programming!

Python code to calculate error metrics

- true_price = [50000, 25000, 10000, 18000,25000, 38000, 40000, 28000,24000]
- model_1_price = [46000, 27000, 9900, 19000, 26000,40000, 43000, 24000, 23000]
- model_2_price = [45000, 26000, 12000, 19000, 27000, 35000, 36000, 29000, 27000]
- print("model 1 mae is:", MAE(true_price, model_1_price))
- print("model 2 mae is:", MAE(true_price, model_2_price))
- print("oracle model mae is:", MAE(true_price, oracle_price))

Classification

- Predict categorical class labels based on past observations
- Class labels are discrete unordered values
- Email spam classification example (binary)
- Handwritten digit classification example (multi-class)



Classification

- Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).
  - The output variables are often called labels or categories. The mapping function predicts the class or category for a given observation.
    - an email of text can be classified as belonging to one of two classes: “spam“and “*not spam*“.
- A classification problem requires that examples be classified into one of two or more classes.
- A classification can have real-valued or discrete input variables.
  - For example, to predict if a fruit is an apple or orange, I will use color as the input variable.
  - To predict if an animal is a dog or cat, I will use weight as the input variable.
- A problem with two classes is often called a two-class or binary classification problem.
  - Dog OR cat?
- A problem with more than two classes is often called a multi-class classification problem.
  - Apple or orange or banana?
  - A problem where an example is assigned multiple classes is called a multi-label classification problem.

Classification evaluation

- Classification evaluation
  - A prediction cannot be perfect
    - There are always errors
    - How to evaluate performance of ML?
  - Evaluation metric: if the prediction is accurate?

Confusion Matrix

- A specific table layout that allows visualization of the performance of an algorithm
  - Each row of the matrix represents the instances in a predicted class
  - Each column represents the instances in an actual class
  - For example:

		Actual class	
		Cat	Dog
Predicted class	Cat	5	2
	Dog	3	3

  - How many animals in this dataset? #dogs? #cats?
  - Now, we can calculate different metrics based on this confusion matrix