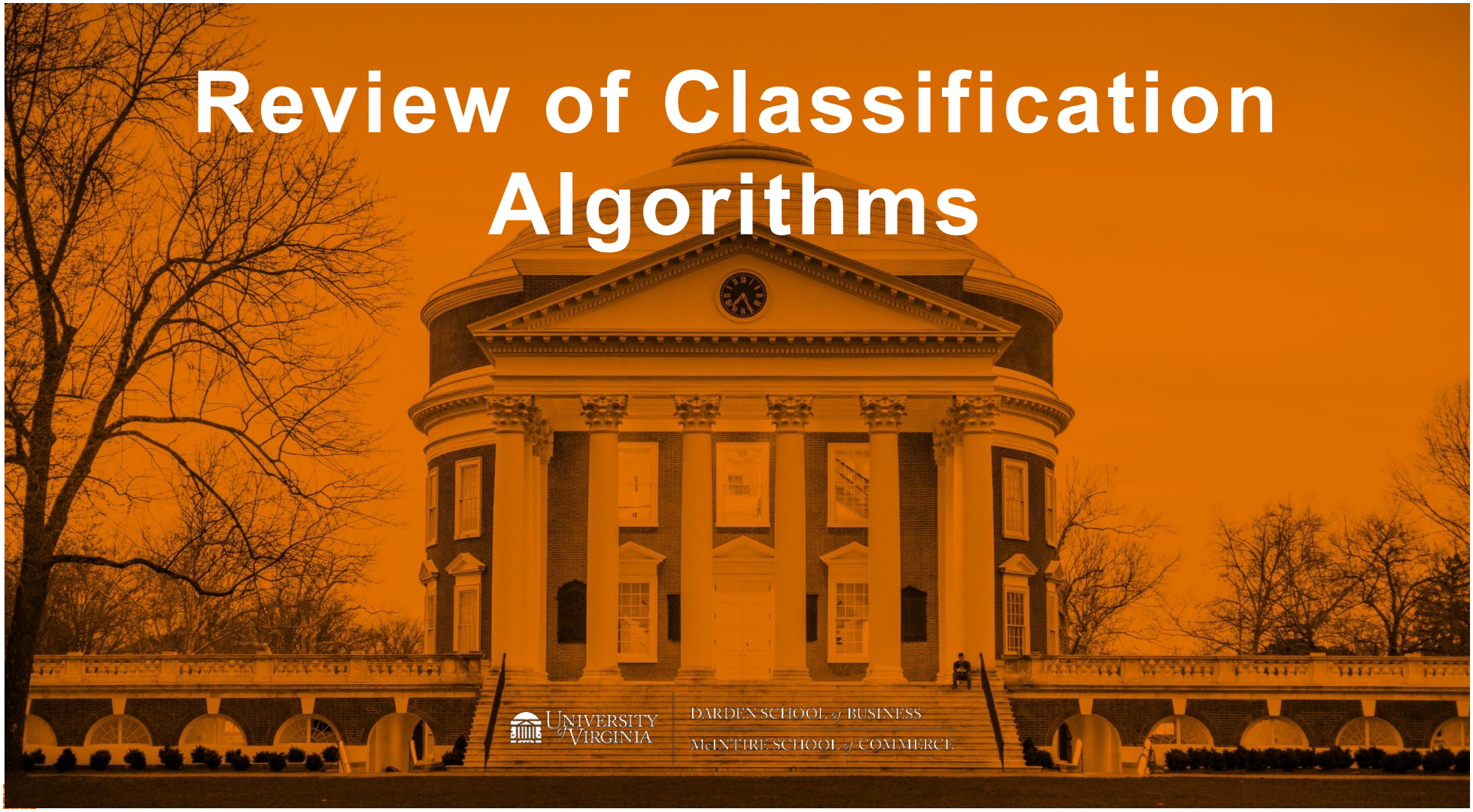
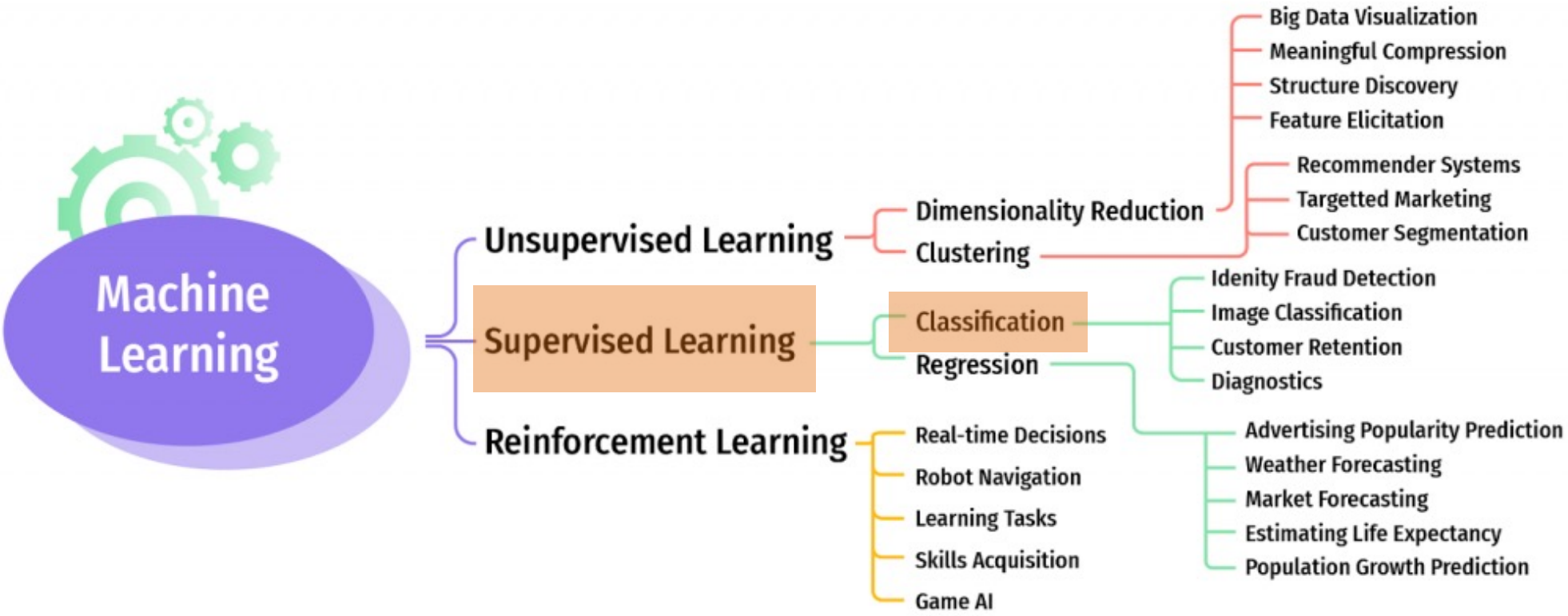
Review of Classification Algorithms




Machine Learning
Unsupervised Learning
Dimensionality Reduction
Big Data Visualization
Meaningful Compression
Structure Discovery
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Idenity Fraud Detection
Image Classification
Customer Retention
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating Life Expectancy
Population Growth Prediction
Reinforcement Learning
Real-time Decisions
Robot Navigation
Learning Tasks
Skills Acquisition
Game AI

The Objective Function in Classification (Cross Entropy)

- :

Cross-entropy is a measure of the difference between two probability distributions for a given random variable or set of events.

ID	Actual	Predicted probabilities
ID6	1	0.94
ID1	1	0.90
ID7	1	0.78
ID8	0	0.56
ID2	0	0.51
ID3	1	0.47
ID4	1	0.32
ID5	0	0.10

ID	Actual	Predicted probabilities	Corrected Probabilities
ID6	1	0.94	0.94
ID1	1	0.90	0.90
ID7	1	0.78	0.78
ID8	0	**0.56**	**0.44**
ID2	0	**0.51**	**0.49**
ID3	1	0.47	0.47
ID4	1	0.32	0.32
ID5	0	**0.10**	**0.90**

The Objective Function in Classification

Now, we take the log of the corrected probabilities (Log(Corrected probabilities)):

ID	Actual	Predicted probabilities	Corrected Probabilities	Log
ID6	1	0.94	0.94	-0.0268721464
ID1	1	0.90	0.90	-0.0457574906
ID7	1	0.78	0.78	-0.1079053973
ID8	0	0.56	0.44	-0.3565473235
ID2	0	0.51	0.49	-0.30980392
ID3	1	0.47	0.47	-0.3279021421
ID4	1	0.32	0.32	-0.4948500217
ID5	0	0.10	0.90	-0.0457574906

Log Loss

The negative average of corrected probabilities is the Log loss or Binary cross-entropy:

$$-\frac{1}{N}\sum_{i=1}^{N}(\log(p_i))$$

ID	Actual	Predicted probabilities	Corrected Probabilities	Log
ID6	1	0.94	0.94	-0.0268721464
ID1	1	0.90	0.90	-0.0457574906
ID7	1	0.78	0.78	-0.1079053973
ID8	0	0.56	0.44	-0.3565473235
ID2	0	0.51	0.49	-0.30980392
ID3	1	0.47	0.47	-0.3279021421
ID4	1	0.32	0.32	-0.4948500217
ID5	0	0.10	0.90	-0.0457574906

Take the average and multiply by -

Log Loss

Log Loss Example- Model A

Observation	Actual Training Label/Class	Prediction Score	Corrected Prediction Score	Log(Corrected Prediction Score)
1	Yes	0.9	0.9	-0.046
2	No	0.2	0.8	-0.097
3	Yes	0.8	0.8	-0.097
4	Yes	0.1	0.1	-1
5	Yes	0.7	0.7	-0.155
6	Yes	0.3	0.3	-0.522
7	No	0.4	0.6	-0.222
8	No	0.3	0.7	-0.155
9	No	0.4	0.6	-0.222
10	No	0.6	0.4	-0.398

Log Loss = 0.291

Log Loss Exercise- Model B

Calculate Log Loss for Model B and compare with that of Model A. Which model is better?

Observation	Actual Training Label/Class	Prediction Score	Corrected Prediction Score	Log(Corrected Prediction Score)
1	Yes	1.0		
2	No	0.1		
3	Yes	0.9		
4	Yes	0.4		
5	Yes	0.8		
6	Yes	0.4		
7	No	0.3		
8	No	0.2		
9	No	0.3		
10	No	0.5		


VIRGINIA