Final Review

Data Science: Jordan Boyd-Graber
University of Maryland
MAY 8, 2018

Normal Distribution

IQ scores are calibrated so that the mean score is 100 and the standard deviation is 15. How many Americans, assuming the US population is approximately 317.4 million, have an IQ greater than 115?

Logistic Regression

Logistic regression classification model that distinguishes American ($Y = 1$) from British documents ($Y = 0$).

| feature	symbol	weight
bias	w_0	1.0
"trousers"	w_1	-1.0
"color"	w_2	2.0
"kerb"	w_3	-0.5
"elevator"	w_4	1.0

1. How is an empty document classified?
2. What is a document where classes are equally balanced?
3. How would "trousers, color, kerb, kerb" be classified?

Suppose we clustered a set of N data points using two different clustering algorithms: *k*-means and Gaussian mixtures. In both cases we obtained 5 clusters and in both cases the centers of the clusters are exactly the same. Can 3 points that are assigned to different clusters in the *k*-means solution be assigned to the same cluster in the Gaussian mixture solution? If no, explain. If so, sketch an example or explain in 1-2 sentences.

A   5

B   -1   Error: -1

C   0.8

D   .6

E   1   Error: 0.2

F   -2
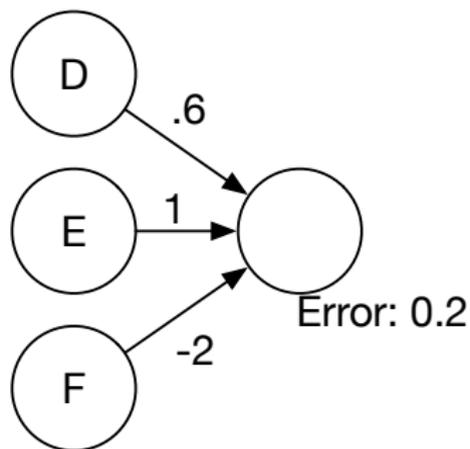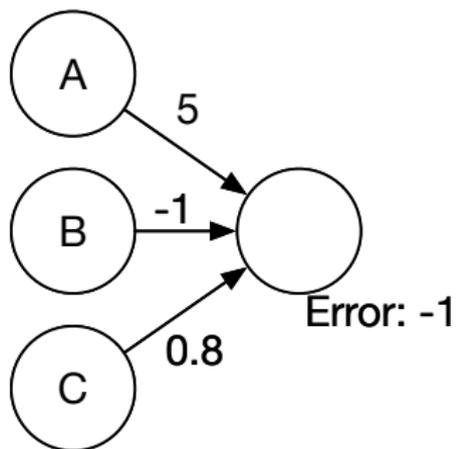
Which node will have the highest backpropagated error?