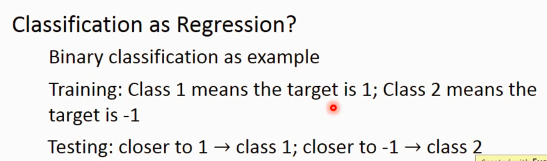
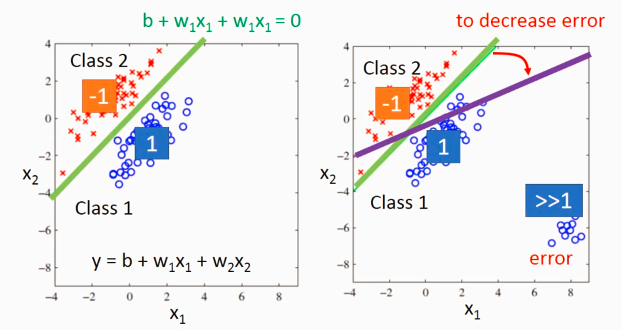
Classification

Caught a cold, the pills made me miss my bed so much, tried my best to keep my upper eyelid away from lower eyelid. What’s worse? There’s a lot of math in this two classes, but my brain asked me to skip it and go back for math when she’s clear.

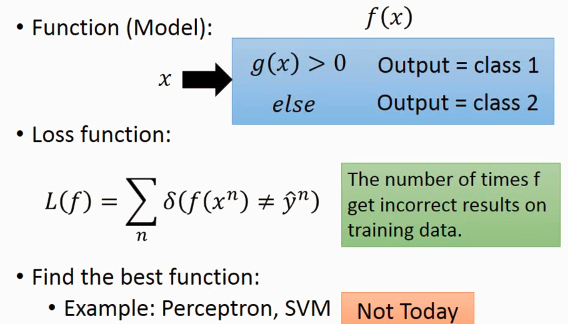
Supposing we want to do classification using regression, like this:



If we put it to 2D coordinate system. The green line is the best estimate boundary for classification while the purple one is the result using regression. Whys that? Because there are some points that are way “too good/true/positive” and it will rotate the final boundary. So regression is not good to classify those data.

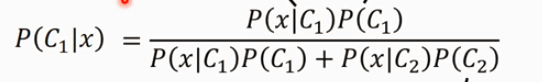


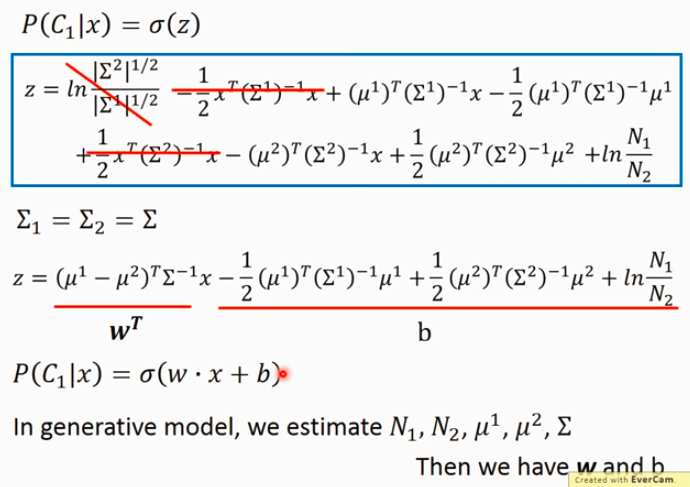
So the ideal alternative is that

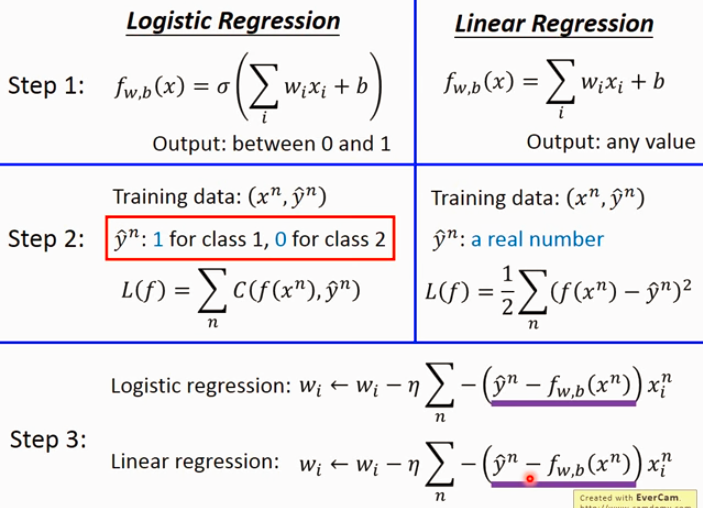


Supposing we have lots of Pokemon, and want to classify them. x is pikachu, C1 the class 1(e.g. water, fire), P(C1|x) presents how likely this pikachu is classified as water Pokemon. The corresponding class with biggest probability is the class pikachu belongs to.

Simply, we need to calculate P(c1|x), P(c2|x), P(c3|x) …, and find the biggest one, then that class is the class of this x.



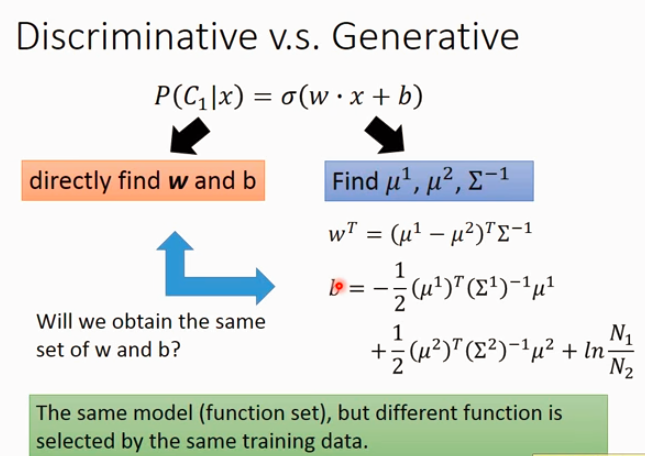


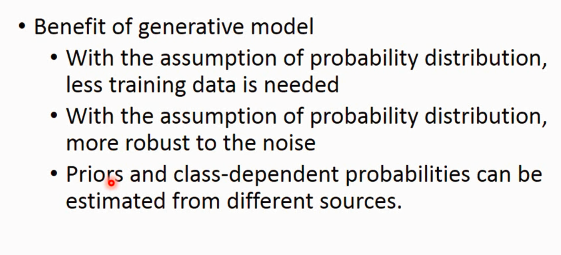


# Discriminative & Generative model

In discriminative model(logistic regression method), there’s no assumptions, but in generative model, we assume the samples conform to some distributions like Gaussian distribution.

Normally, the accuracy of discriminative model is higher than generative model. Because in generative model, it is assume that the all features are independent, but in practice, they are not totally independent.





Understanding for the points above:

1. The number of training data have more influence on discriminative model than on generative mode. When the amount of training data is few, since the generative model has its own assumption, so it would follow its heart then follow the given data.
2. Again, some times the features would be noisy, but since generative model have its assumption, it would follow its heart, and ignore the noise.
3. E.g. if we grab lots of articles online, we can have the probabilities distribution.

<http://scikit-learn.org/stable/modules/preprocessing.html#feature-binarization>