Deep Learning Types

Generative Adversarial Networks (GANs)

Intro

A Generative Adversarial Network (GAN) is not one network, but rather a pair of distinct networks, a Generator and a Discriminator. The Generator seeks to generate false samples indistinguishable from real samples. The Discriminator seeks to discriminate between the false samples and the real samples. By training and evolving alongside each other, they both learn their task better.
The theoretical end result is two distinct and useful networks: a Generative network that can generate samples indistinguishable from training data, as well as a Discriminative network that can detect the faintest hint of a false sample.

Key Vocabulary / Abbreviations

- Generative Adversarial Network (GAN) - the whole system
- Generative Network (G) - the half of the system that generates false samples
- Discriminative Network (D) - the half of the system that evaluates samples for legitimacy

Links:

- 2017 Computerphile video about GAN basics: GAN

Variational Autoencoders (VAE)

Intro

Links:

- 2018 Video with intuitive explanation: Autoencoders

Logistic Regression

Intro

We stick a normal linear model into a sigmoid.

Explanation

Sometimes we want to use a simple linear model to calculate a probability. Perhaps as the temperature of a star gets arbitrarily cold, its odds of collapsing increase. But also its odds go up as its size decreases or as its fuel supply runs out. We could (maybe inaccurately) model the probability of collapsing using a weighted linear regression, but that easily gives values outside 0 to 1. So we sigmoid it, to ensure we have reasonable probability values.

I believe (?) that Logistic Regression always results in inputs that follow a normal distribution given a specific output.

Gaussian Descriminant Analysis (GDA)

Intro

Broadly speaking, this is a generative method which seeks to model a continuous input space given an output. The key piece of this analysis is that it **assumes that the input space for a given output follows a multivariate normal distribution** (this is basically just a normal distribution across each dimension).

Intuitive explanation

The model first subsets the data based on the output values. It then models each of these subsets using a multivariate normal distribution (a higher-dimensional normal distribution). To generate a new sample output, it samples a point from the appropriate distribution. Hopefully the mean values of each distribution are different, or else the model will have no power to classify, only to generate (and that probably wouldn't work either depending on why the means are the same).
The assumption that the model is multivariate normally distributed is a big assumption. If it is correct, then GDA is asymptotically efficient (as data set gets large) and easy to train. If this assumption is wrong, presumably garbage.
Logistic regression is similar to GDA and much more robust (likely better in most cases).

Links

- CS Lecture on GDA and Naive Bayes by Andrew Ng GDA and Naive Bayes

Naive Bayes

Broadly Naive Bayes takes a set of discrete inputs and finds a probability distribution for each input. It combines this with a discrete output and gives probabilities for each input value given the output.
For example, in a house it might take the number of bathrooms, number of bedrooms, color paint, and number of stories, and the city which the house is in. Then given a city, it will give the probability of each other variable.
Critically, this **assumes that given a certain output, the value of each input is independent of each other**. This is ridiculous in the house example: if I know a house is in Seattle, the number of bedrooms still affects my predicted number of bathrooms. For that reason this model is almost never useful.

CS Lecture on GDA and Naive Bayes by Andrew Ng GDA and Naive Bayes

Other Links:

- 2018 Blog post about Normalizing Flows: Normalizing Flows
- Distill - website with articles about Machine Learning: Distill Site