1 Studying Gaussian Processes

1.1 References

1. Distill.pub
2. Introduction to probability(by Berstekas & Tsitsiklis)

1.2 Covariance and Correlation

These two *measure* the *strength* and *direction* of the relationship between two random variable. By el way : these are used for *Bayesian Statistical Inference* and *Classical Statistical Inference*.

Covariance of Two random Variables X and Y: It is denoted by

$$cov(X.Y) = \mathbf{E}[(x - E[X])(Y - E[Y])]$$

When cov(X,Y) = 0 we say that X and Y are **uncorrelated**.

Stochastic Process A mathematical model that evolves in time and *generates a sequence of numerical values.* Sequences can be something like daily prices of stock(im assuming each value representing a price at a day). *Each of these values is modeled by a random variable.* DO keep in mind that all random variables in a stochastic process refer to a single and *common* experiment, and are thus defined in a common sample space.

In stochastic processes we focus on :

- **Dependencies** in the squence of values generated. e.g. how does the value of predicted stock prices *depend* on past values.
- We focus on **long term averages** involving the *entire sequence* of generated values.
- We sometimes wish to know the frequency of certain **boundary events** (extreme cases).

Markov Processes In this case we work with experiments that evolve in time and in which the future evolution exhibits a **probabilistic dependence** on the past. e.g. If we wish to "predict" stock prices we have to keep in mind that they depend on the prices of the past.

Guassian Processes Tools that allow us to make prediction of our data by taking into account prior knowledge. Remember that there are an infinite amount of functions that can fit collected data. **Gaussian Processes help us to assign a probability to each of these functions. It is easy to see then that the mean of this probability distribution represents the most probable characterization of the data**

For this reason they can be applied to regression. Gaussian Processes are not limited to regression though! THey can also work for classification and clustering.

The Gaussian distribution is the building block of Gaussian processes. We are however interested in the multivariate one where each variable has a Gaussian distribution and their joint distribution is *also* a Gaussian distribution

As expected the *Gaussian Distribution* is defined by a mean vector μ(one per dimension representing the dimensions mean) and a covariance matrix Σ

These processes are closed under both *conditioning* and *marginalization*. This means that the resulting distributions from these operations are also Gaussian. **This makes many problems in statistics and machine learning tractable**.

//The need for a small learnign rate is due to the variance of SGD(SGD approximates the actual gradient using batches. *This introduces Variance*)

2 Back to the VOG paper

2.1 Reference

1. Agarwal & Hooker

2.2 Notes

1. Again this is about estimating difficult examples for a model to classify.
2. Data points with high VOG scores are far more dfficult for the model to learn and over-index on corruped or memorized exmaples.

What if we can get a set of images that are high in VOG and we sparsify them with more care.