

Saddle Points Continued, Max min principle

We will begin with the saddle points. The reason why we interested in saddle points is when we get to this deep learning direction, the big step these is finding a minimum of the total cost function and gradient descent, which we will discuss as the usual method or stochastic gradient descent. And all kins of issues arise, what happens if you have a saddle point or a degenerate minimum? All these possibilities and the understanding of deep learning is focusing more and more on what does that the gradient descent algorithm produce.

What we want to know is what is the maximum and minimum of the Rayleigh quotient

$$R(x) = \frac{x^T S x}{x^T x}$$

And what is the saddle point? the middle saddle value. is the maximum of the minimum or the minimum of the maximum of one subspace.

Sample mean, expected mean.

Sample variance, variance, Covariance.