Computational Neuroscience
Dr. Sharba Bandyopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology Kharagpur
Week – 08
Lecture – 36

Lecture 36: Statistical Methods in Discrimination

Welcome. So, we will switch gears a little bit here and go on to some of the statistical approaches that are there in terms of decoding response decoding the value of a stimulus or what the stimulus was based on responses. And so we will treat this in the case of so we have been talking about discrimination and for the single neuron case as well as multiple neuron cases. However, we have been treating stimuli as discrete elements that is mostly that we have a set of stimuli 1 to capital N and then we are we created let us say the confusion matrix or we created how I mean based on spike train distances we also did look at population based approaches we looked at KL distance and so on. However, other than decoding a stimulus over time let us say when we talked about S(t) and basically finding out estimating S(t) from the response over time we treated the stimuli as nonparametric. Now and in these cases where we have multiple stimuli it essentially means that the stimuli are very different from each other even if they are parametric they are very different from each other.

However, we may have situations where we may need to deal with very fine changes in the stimulus and these are applicable to the case of parametric stimuli mainly whatever we are going to discuss in these in this part of the course. So let us say that we have a stimulus S which is a particular value a scalar or rather a variable which takes on one value for now and this may be changing only slightly to let us say S plus delta S and by looking at various features of the response for such small changes in stimuli we will be able to see how well the responses or that neuron encodes the parameter S. So that means if for a small change there is a large change in the response then it would mean that there is a high fidelity of encoding the stimulus S. So in the case that we will be studying that okay for a stimulus we have a set of responses and from the responses we are going to estimate the stimuli or the stimulus S estimate that particular value of the parameter of the stimulus.

So think of S as a parameter here and response is the random variable based on which or the observations based on which we are going to make an estimate about S. So an important factor here or important term here is what we call the score which is nothing but the derivative of log of response given a stimulus S with respect to S. So the partial derivative I am sorry log of probability of response to the stimulus S with respect to S. So essentially this term here is what we call the log likelihood that is what is the likelihood of the response being some R given a stimulus logarithm of that and how sensitive is that likelihood to changes in S or the stimulus. So the more the sensitivity the more better will be the fidelity of encoding by the response and so based on this we will based on this idea we will go further forward and with this score will come back up again in our discussions.

So we will be dealing with this probability distribution P R given S that is what an experimenter gets so this is results of experiments this is our observations or data or observations or data. So with multiple repetitions we get a distribution of R for a given stimulus S or stimulus parameter S. So let us continue with this idea in a slightly different manner. So we have R given S that is given a particular stimulus with multiple repetitions we are getting a probability distribution of R given S. From that we are estimating S estimate and that let us say is on average for multiple when we if we were to able to do this multiple number of times then let us say that it is mean it has a mean that is same as the true value of the parameter S.

Now in very different in multiple different cases if with the same data if we estimate multiple times we will have different values of estimates if we were to do the experiment again and again in different cases we may have the estimates to be spread around this true value of S. Now earlier we were talking about the score which is our log of P of r given S the derivative of this with respect to S. This score says that for a particular distribution of response given a stimulus how the log likelihood of the response changes with small changes in the stimulus that is how sensitive the response is to changes in stimuli. If we have larger changes that would mean that it is coding in a better way. However here we will change the idea slightly and talk about it in a different manner in the sense that we will say that any estimator S estimate any estimator has an associated variance let us call that sigma square sorry let us call that $sigma^2$ estimate that is the variance of the estimator.

Let us now we are considering unbiased estimator and so there is variability in the estimates in the sense that we can have an based on the data we can have an estimator that has a distribution of the estimates for multiple different experiments that is narrow very narrow very close to the true value of S. We can also have an estimator whose variance is high with a large spread of the estimates around the true value S which means that the one that is narrow that is a better estimator that is the variance is low. So, always an estimator we will be trying that the estimator has low variance that is it is as close as possible to the true value S for different cases different experiments that is how we would like the things to be we would not want because we can be anywhere on this line in within this yellow range for a given experiment is what we are saying and so the chances of being wrong is high. Similarly if we have the many the variance of the estimator to be low the chances of being very much away is extremely low. So, this way we can define the fidelity of coding of the stimulus based on the responses.

So, this whole thing is being determined from P r given S this P r given S. So, there is a minimum variance that is associated with an estimator that is given by the Cramer Rao bound that is this $sigma^2$ estimate that is the variance in the estimator is greater than equal to 1 over f S. So, this is also dependent on S that is the Fisher information this is called the Fisher information. So, the Fisher information is actually we can let us put a scalar here in any case just. So, if we have so the Fisher information is given as the expectation of the square of the score that is del log P r given S del S squared.

So, this expectation is obviously on r that the response distribution. Alternatively this is an exercise that you can do this can also be defined as expectation of the negative second partial derivative of log P r given S del S. So, what we are saying here that if this f S is low Fisher information is low that is based on the data that we have collected that is P r given S we estimate the Fisher information f S for that particular stimulus at that particular location at that for that parameter. What the data is telling us in that case if f S is low is that the minimum variance of the estimator that we can reach or the best case scenario that can be reached based on the data is quite high variation or high variance. So, in other words there will be chances of it being like the yellow distribution the different estimates in if we were to repeat the experiment many many times we will get a large variation.

So, what with the low Fisher information what we are seeing is that the best possible scenario is going to be poorer than the case where the Fisher information is high that if the Fisher information is high then the minimum variance of an unbiased estimator this $sigma^2$ estimate that can be reached is much lower than the case where the Fisher information is low. So, based on the Fisher information then we can conclude about how well the response encodes the stimulus or in other words if we are trying to estimate the stimulus from the response how well we can estimate that is determined by the Fisher information which is as we have shown is this expectation of the square of the score. So, as we will see as you can think about it a little more. So, cases where we have the response variability to be low as well as changes of the responses with stimulus is high this combination is going to produce the highest Fisher information. If the variance is constant throughout the stimulus change space then it is the change in response with the stimulus that is the most important thing.

If the response change is moderate or I mean is more or less constant with the stimulus change then where we have the least variability in the responses for that stimulus that is where we will have the highest fidelity of coding of that particular stimulus parameter. So, with this we will conclude our first discussion about statistical approaches in estimating stimuli from responses. We will continue this discussion in the next section. Thank you.