Week 3 Lecture Notes

Points of clarification and fun facts re: video lectures

L1: Neural Decoding and Signal Detection Theory

- Slide: Do I Stay or Do I Go?
    - In this graph, the x-axis represents some stimulus parameter (such as the amplitude of the noise), and the y-axis represents probability (density). Noises made by the tiger are represented by the red distribution and noises made by the wind are represented by the blue distribution. If the x-axis is amplitude, the red curve being to the right of the blue curve indicate that on average, the tiger is going to make a louder noise than the wind. However, this isn't always the case, as is indicated by the overlap in the two curves.
    - In neuroscience, the x-axis might represent the activity of some neuron (or the average activity of a group of neurons), which the rest of your brain has to readout in order to make a decision
- Slide: Predictable from neural activity?
    - In the histogram, the x-axis is the activity of the neuron and the y-axis is the number of counts (the number of trials in which the neuron had a particular response)
- Slide: Signal detection theory
    - A false alarm is equivalent to choosing the answer upward when the true answer was downward; this is equivalent to the probability that the downward stimulus leads to a firing rate above the threshold z (because everything above the threshold z you label as upward, according to your rule), i.e., the probability of the firing rate being above z *given* that the stimulus was downward: $P(r > z \mid \text{downward})$.
- Slide: Likelihood ratio
    - The likelihood of a model (e.g., the stimulus being upward or downward moving) is equal to the probability of seeing the data (e.g., the firing rate) *given* that model. This is *not* the same as the probability of the model *given* the data, although the two are related through Bayes' rule.
- Slide: Let's just consider for a moment
    - When we say that making many observations leads to multiplying many probabilities together, we are assuming that the observations are independent across time, meaning that $P(s1, s2|\text{tiger}) = P(s1|\text{tiger})P(s2|\text{tiger})$, where s1 and s2 are observations at two different times. Even if s1|tiger and s2|tiger are not independent, calculating probabilities of a long sequence of observations can lead to numerical errors in computation (for example, MATLAB may round very low probabilities to zero). Because of this, we often work with logarithms, which transform products into sums, allowing us to avoid these errors.
- Slide: Nonlinear separation of signal and noise

- In the plots of P(I|signal) and P(I|noise), the y-axis follows a logarithmic scale. This means that Gaussian probability distributions will appear as inverted parabolas. Can you think of why this is?
- Prior probabilities are very important in decision making, and they can be expressed mathematically using the chain rule of probabilities. Bayes' rule tells us, for example, that P(tiger|sound) is proportional to P(sound|tiger)P(tiger). We don't need to worry about the proportionality constant if we are only comparing P(breeze|sound) to P(tiger|sound), which we usually are.

- Slide: Building in cost
  - In other words, the three important pieces of information you need when making a binary decision are: the evidence – the value of the sound, the prior – P(tiger), (the two combine to give the posterior P(tiger|sound)), and the cost/loss associated with a wrong/right decision.

L2: Population coding and Bayesian estimation

- Slide: Cricket cercal cells
  - The four humps on the firing rate/degree plot represent tuning curves for the 4 different types of neuron.
  - The cosine-like response functions of the cells have been shifted and scaled so that the normalized firing rate is always between 0 and 1.
- Slide: Population coding in M1
  - The eight different clumps of vectors represent eight different neurons, each of whom has a preferred direction indicated by its position among the other seven neurons.
- Slide: Decoding an arbitrary continuous stimulus
  - The Gaussians here are not probability distributions. "Gaussian" just refers to the mathematical form of the tuning curve.
- Slide: Maximum likelihood
  - Notice the form of this probability distribution. P[**r**|s] takes as input a vector of responses (a list of several neurons' responses) gives as output a single number, a probability.
  - Assuming you've measured/observed a real response **r**, ln(P[**r**|s]) is a function only of s. Therefore, it can be maximized with respect to s.
- Slide: Maximum likelihood: setting the derivative equal to zero
  - This means that our answer for s* is the most likely stimulus, given the data (**r**). Strictly, you would also have to show the second derivative of ln(P[**r**|s]), with respect to s, is negative (so that we are sure we have a maximum), but this hasn't been done in the lecture.

L3: Stimulus reconstruction

- Slide: The role of the conditional mean
  - In general if you have a random variable x, the average/expected value of some function of x, f(x), is equal to $\int f(x)p(x)dx$. Since one integrates over x, this expression no longer becomes a function of x.

Similarly, the expected error is only a function of the estimator, s_{Bayes}, and the recorded firing pattern, **r**. If the firing pattern has been fixed (i.e., recorded such that it can no longer change), then the expected error is only a function of s_{Bayes}. Thus, it makes sense to choose s_{Bayes} such that the expected error is minimized.

  - When working with an expression involving an integral, remember that the expression is not a function of the variable that is integrated over, e.g., $\int sp[s|\mathbf{r}]ds$ is *not* a function of s; it is only a function of **r**.
  - Note that $\int sp[s|\mathbf{r}]ds$ is the conditional mean, i.e., the expected value of s given the firing rates **r**. This is very intuitive, as it just implies that if you measured some firing rates **r** and are trying to figure out the stimulus, your best guess would be what you expect it to be, given the data (the firing rates) you've observed.
- Slide: Reading Minds: fMRI
  - In this case, both **r** and s are very high dimensional (so maybe it would be better to write s as a vector **s**). **s** has an element for every pixel value at every point in time in the movie. What do the elements of **r** correspond to?
  - P(s) being uniform across samples just means that each sample is assigned an equal probability. If you were to draw this as a continuous function (i.e., a probability density), there would be a scaled Dirac delta function at every sample in the true set; everywhere else would be zero.