BioMath Bootcamp - Model Fitting Problem Set

(1) Poisson Regression

- Write code to implement the Poisson regression example shown in class. Pick three positive stimulus values x_1, x_2, x_3. Let $\theta = 2$, and generate spike counts $y \sim Poisson(\theta x)$. Draw five responses y for each stimulus value. The matlab function poissrnd will be useful.
- Plot the likelihood function $P(y|x, \theta)$, as a function of θ. Also plot the log-likelihood function.
- Compute the analytic maximum likelihood estimator for θ. How close is it to the true θ?
- Compare the poisson MLE to the MLE under gaussian noise. Which one has lower error? Repeat the experiment 100 times and compute the standard deviation of the errors of each estimator.

(2) Exponential Regression

Suppose we have a neuron that responds linearly to a stimulus x with a slope parameter θ (just like before), but now the noise is controlled by an exponential distribution:

$$P(y|x) = (\theta x)e^{-(\theta x)y}$$

- Derive the maximum likelihood estimator for θ based on a set of x_i and y_i data points (just like before). You should be able to find a closed form expressions.

(3) Logistic Regression

In binary classification, each datapoint x is mapped to one of two possible labels: $y = 0$ or $y = 1$. Logistic regression maps this problem into a regression setting by modeling the continuous probability that $y = 1$ as a function of independent regressors x.

Logistic regression is an example of a generalized linear model (note: this is not the same as a general linear model), in which the nonlinearity is the logistic function, 1/(1+ exp(-x)), and the noise distribution is bernoulli "coin flipping." The simplest case has one parameter, slope θ. The logistic regression model is:

$$p(y = 1|x) = \frac{1}{1 + \exp(-\theta x)}$$

$$p(y = 0|x) = 1 - p(y = 1|x) = \frac{1}{1 + \exp(\theta x)}$$

- Given a dataset of pairs x_i, y_i, where $x_i \in \mathbb{R}$, and $y_i \in \{0, 1\}$, show that we can write the likelihood as:

$$P(y|x, \theta) = \prod_i \frac{\exp(\theta x_i y_i)}{1 + \exp(\theta x_i)}$$

- Show that if x_i are not all identical, then there is no analytic expression for the MLE of θ.
- Draw a simulated dataset with $\theta = 1.5$. Sample 20 values x_i from a standard normal distribution. For each x_i, sample y_i according to the formula for $p(y{=}1|x)$ above. That is, y_i is either 0 or 1, with a probability that depends on x_i according to the formula above. Scatter plot the raw data.
- Plot the likelihood and log-likelihood functions for $\theta = -5$ to 5. Is the maximum close to the true value of θ?
- Draw more samples. How does the width of the likelihood change as you add more samples? Based on your observations, what does the width of the likelihood function tell you?
- Use matlab's function fminunc to numerically find the MLE (same as just looking at the likelihood function for the maximum point).