STA613/CBB540 HOMEWORK 3

DUE THURSDAY, FEB 7^{TH}

(1) *Poisson regression and iteratively reweighted least squares (IRLS) via Newton Raphson.*
   (a) Derive the iteratively reweighted least squares algorithm for Poisson regression,
   (b) implement IRLS for Poisson regression,
   (c) generate 500 Gaussian data points (see: rnorm) x and, for three different θ values (one negative, one between 0 and 1, and one greater than one) generate 500 poisson-distributed points $y_i, i \in \{1, ..., 500\}$ with mean $\mu_i = \exp \theta x_i$ (see: rpois). Plot these three data sets (try x versus y, x versus log y).
   (d) estimate, using your IRLS code, the θ parameters for x each of the three θ and y. How many iterations does it take to converge to something reasonable? How far away is this estimate from the truth? Show (three on the same plot – see: lines) the value of each θ estimate (y-axis) at each iteration (x-axis).

(2) *Bayesian regression.* In class, we discussed the expected value of the effect size β in a linear model in Bayesian linear regression, given a Gaussian prior on β. We might let this expected value of the effect size be our estimate of β in the Bayesian setting, call this $\tilde{\beta}$. Call the (ML) estimate of β in the frequentist setting $\hat{\beta}$.
   (a) Write code to compute $\tilde{\beta}$ for a given set $D = \{(x_1, y_1), ..., (x_n, y_n)\}$, where x_i are scalar, and μ_0 (the mean of the distribution on β) and λ_0 (the variance of β) (although there is a scalar predictor x, remember to include an intercept term β_0).
   (b) Download genotypes and gene expression data from our earlier homework. Compute $\hat{\beta}$ and $\tilde{\beta}$ for each pair of SNPs and gene expression values for $\mu_0 = 0$, $\lambda_0 = 1$. Plot the values (one on the x-axis, one on the y-axis) and draw a line representing $x = y$. What do you notice about the Bayesian estimate as compared to the frequentist estimates?
   (c) Create the same plot for i) a very large value of λ_0 and ii) a very small value of λ_0 (non-negative). What effect does this value have on the $\tilde{\beta}$?
   (d) Try varying μ_0. What impact do different values of μ_0 have on $\tilde{\beta}$?