Franklin Hu, Sunil Pedapudi
CS 194-10 Machine Learning
Fall 2011
Assignment 5

1. Conjugate priors

   (a)

   (b) Given the geometric distribution

   $$P(X_i = k|\theta) = (1-\theta)^{k-1} \cdot \theta$$

   and the beta distribution

   $$\beta(\theta|a,b) = \alpha\theta^{a-1}(1-\theta)^{b-1}$$

   we prove that the beta distribution is the conjugate prior for a likelihood with a geometric distribution.

   $$\begin{aligned} P(\theta|X) &= P(\theta) \cdot P(X|\theta) \\ &= \alpha \cdot \theta^{a-1} \cdot (1-\theta)^{b-1} \cdot (1-\theta)^{k-1} \cdot \theta \\ &= \alpha \cdot \theta^{a} \cdot (1-\theta)^{b+k-2} \\ &= \beta(\theta|a+1, b+k-1) \end{aligned}$$

   The posterior has the form of a beta distribution so therefore the beta distribution is the conjugate prior for the geometric distribution.
   The update procedure for a beta posterior simply involves updating the a and b parameters

   $$\begin{aligned} a_{N+1} &\leftarrow a_N + 1 \\ b_{N+1} &\leftarrow b_N + k - 1 \end{aligned}$$

2. Bayesian Naive Bayes

3. Logistic regression for credit scoring

   (a) The data structure we chose for logistic regression is simply a class that keeps a set of weights for each of the features, has an update method for updating the weights, and draws predictions using the logit function

   $$\text{Probability} = \frac{1}{1+e^{-w^T x}}$$

   (b) The likelihood is

   $$\begin{aligned} L(w) &= \frac{1}{1+e^{-yw^T x}} \\ \text{log likelihood} &= \log\frac{1}{1+e^{-yw^T x}} \\ &= -\log(1+e^{-yw^T x}) \\ \text{negative log likelihood} &= \log(1+e^{-yw^T x}) \end{aligned}$$

Now we compute the gradient of the negative log likelihood

$$
\begin{aligned}
\nabla \log(1+e^{-yw^Tx}) &= \nabla \log\left(\frac{e^{yw^Tx}+1}{e^{yw^Tx}}\right) \\
&= \nabla\left(\log(e^{yw^Tx}+1) - \log(e^{yw^Tx})\right) \\
&= \left(\frac{1}{e^{yw^Tx}+1}\cdot e^{yw^Tx}\cdot -yx_i\right) - \left(\frac{1}{e^{yw^Tx}}\cdot e^{yw^Tx}\cdot -yx_i\right) \\
&= yx_i - yx_i\cdot\frac{e^{yw^Tx}}{e^{yw^Tx}+1} \\
&= yx_i - yx_i\cdot\left(\frac{e^{yw^Tx}+1}{e^{yw^Tx}}\right)^{-1} \\
&= yx_i - yx_i\cdot(1+e^{-yw^Tx})^{-1} \\
&= yx_i\left(1-\frac{1}{1+e^{-yw^Tx}}\right)
\end{aligned}
$$

Therefore our update rule is simply

$$
\begin{aligned}
w_{i+1} &= w_i + \alpha\cdot\nabla L \\
&= w_i + \alpha\cdot yx_i\cdot\left(1-\frac{1}{1+e^{-yw^Tx}}\right)
\end{aligned}
$$