Homework 2

Due Date: March 30, 2018

Problem 1. The Poisson distribution is a useful discrete distribution which can be used to model the number of occurrences of something per unit time. For example, in networking, packet arrival density is often modeled with the Poisson distribution. If X is Poisson distributed , i.e. X~ Poisson(λ), its probability mass function takes the following form:

$$\mathrm{P}(\mathrm{X}|\lambda) = \frac{\lambda^X e^{-\lambda}}{X!}$$

It can be shown that E(X) = λ. Assume now we have n i.i.d. data points from Poisson(λ): $\mathrm{D} = \{X_1, \ldots, X_n\}$.

(For the purpose of this problem, you can only use the knowledge about the Poisson and Gamma distributions provided in this problem.)

(a) Show that the sample mean $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the maximum likelihood estimate (MLE) of λ and it is unbiased ($\mathrm{E}(\hat{\lambda}) = \lambda$).

(b) Now let's be Bayesian and put a prior distribution over λ. Assuming that λ follows a Gamma distribution with the parameters (α, β), its probability density function:

$$\mathrm{p}(\lambda|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

where Γ(α) = (α−1)! (here we assume α is a positive integer). Compute the posterior distribution over λ.

(c) Derive an analytic expression for the maximum a posterior (MAP) of λ under Gamma(α, β) prior.

Problem 2. Consider the following training set, in which each example has two tertiary attributes (0, 1, or 2) and one of two possible classes (*X* or *Y*).

Example	A1	A2	Class
1	0	1	X
2	2	1	X
3	1	1	X
4	0	2	X
5	1	2	Y
6	2	0	Y

Please answer the following questions with details.

1. What would the Naive Bayes algorithm (with no smoothing) predict for the class of the following new example?

Example	A1	A2	Class
7	2	2	?

2. Apply Laplace smoothing (with a smoothing parameter equal to 1) to all the probabilities from the training set. Will the result of the previous question change?

Problem 3. when Y is Boolean and $\mathrm{X} = < X_1 \ldots X_n >$ is a vector of continuous variables, then the assumptions of the Gaussian Naive Bayes classifier imply that P(Y|X) is given by the logistic function with appropriate parameters W. In particular:

$$\mathrm{P}(\mathrm{Y} = 1|\mathrm{X}) = \frac{1}{1 + \exp(w_0 + \sum_{i=1}^{n} w_i X_i)}$$

and

$$\mathrm{P}(\mathrm{Y} = 0|\mathrm{X}) = \frac{\exp(w_0 + \sum_{i=1}^{n} w_i X_i)}{1 + \exp(w_0 + \sum_{i=1}^{n} w_i X_i)}$$

Consider instead the case where Y is Boolean and $\mathrm{X} = < X_1 \ldots X_n >$ is a vector of Boolean variables. Prove for this case also that P(Y|X) follows this same form (and hence that Logistic Regression is also the discriminative counterpart to a Naive Bayes generative classifier over Boolean features).

Hints:

• Simple notation will help. Since the X_i are Boolean variables, you need only one parameter to define$\mathrm{P}(X_i|\mathrm{Y} = y_k)$. Define $\theta_{i1} = P(X_i = 1|Y = 1)$, in which case $P(X_i = 0|Y = 1) = 1 - \theta_{i1}$. Similarly, use θ_{i0} to denote $P(X_i = 1|Y = 0)$.

• Notice with the above notation you can represent $P(X_i|Y = 1)$ as follows

$$P(X_i|Y = 1) = \theta_{i1}^{X_i}(1 - \theta_{i1})^{(1 - X_i)}$$

Note when X_i = 1 the second term is equal to 1 because its exponent is zero. Similarly, when X_i= 0 the first term is equal to 1 because its exponent is zero.