CS5350: Machine Learning — *Due 18 Nov 2008*

HW8: Probabilistic Modeling

1 Written Exercises

1. Suppose we have a binomial distribution with the "probability of heads" $\pi = 0.8$. Compute (show all the stops) the expected value and variance of this distribution.

   Answer: the probability distribution over x can be written as:

   $$\mathcal{P}(x|\pi) = \pi^x (1-\pi)^{1-x}$$

   The expectation of x is

   $$E(x) = \sum_{x=0,1} p(x)f(x) = \pi \cdot 1 + (1-\pi) \cdot 0 = \pi$$

   The variance of x is

   $$var(x) = E[x - E(x)]^2 = E(x^2 - 2x \cdot E(x) + E^2(x)] = E(x^2) - E^2(x)$$

   Because

   $$E(x^2) = \sum_{x=0,1} p(x)x^2 = (1-\pi) \cdot 0^2 + \pi \cdot 1^2 = \pi$$

   We have

   $$var(x) = \pi - \pi^2 = \pi(1-\pi)$$

2. Suppose we have a Gaussian with known mean $\mu = 1$ and known variance $\sigma^2 = 1$. What is the *density* of the distribution $\mathcal{N}or(\mu, \sigma^2)$ at the following points: $0, 1, 2$?

   Answer: the Gaussian distribution function with $\mu = 1$ and $\sigma^2 = 1$ can be written as

   $$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}$$

   Hence,

   $$\mathcal{N}(0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}} = 0.2420, \mathcal{N}(1) = \frac{1}{\sqrt{2\pi}} = 0.3989, \mathcal{N}(2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}} = 0.2420$$

3. Consider the previous question, but where $\sigma^2 = 0.1$. What is the density at the given points? For $x = 1$, the density should be *greater than one*. How is this possible given that the Gaussian is normalized (i.e., sums to one).

   Answer: when $\sigma^2 = 0.1$, the distribution function is

   $$\mathcal{N}(x) = \frac{1}{\sqrt{0.02\pi}} e^{-50(x-1)^2}$$

   And

   $$\mathcal{N}(0) = 7.6946e^{-22}, \mathcal{N}(1) = 3.9894, \mathcal{N}(2) = 7.6946e^{-22}$$

   The $\mathcal{N}(1) = 1$ means at point $x = 1$, the probability of x falls into the interval $(x, x + \delta x)$ is $\mathcal{N}(x) \cdot \delta x$ when $\delta x \to 0$. So, even $\mathcal{N}(x) > 1$ at this point, the $\mathcal{N}(x) \cdot \delta x$ is less than 1.

4. The *Poisson* distribution is a distribution over *positive count values*. It has the form $p(k \mid \lambda) = \frac{1}{e^\lambda} \frac{\lambda^k}{k!}$, where k is the count and λ is the (single) parameter of the Poisson. Suppose we have a bunch of count data (for instance, the number of cars to pass an intersection on a given day, measured on N-many days) called $k_1, k_2, \ldots, k_N$. Compute the maximum likelihood estimate for λ given this data. (Hint: write down the likelihood, then take the log. Do some algebra to simplify and then take the derivative with respect to λ.)

   Answer: The likelihood function of λ is

$$P(data|model) = P(k_1, k_2, \cdots k_N|\lambda) = \prod_{i=1}^{N} P(k_i|\lambda)$$

$$= \prod_{i=1}^{N} \frac{1}{e^\lambda} \cdot \frac{\lambda^{k_i}}{k_i!} = e^{-\lambda N} \prod_{i=1}^{N} \frac{\lambda^{k_i}}{k_i!}$$

   And the log-likelihood of λ is

$$L(\lambda) = -N\lambda + \sum_{i=1}^{N} \log \frac{\lambda^{k_i}}{k_i!} = -N\lambda + \sum_{i=1}^{N} k_i \log \lambda - \sum_{i=1}^{N} \log k_i$$

   Take derivative with respect to λ, we get

$$\frac{\partial L(\lambda)}{\partial \lambda} = -N + \sum_{i=1}^{N} k_i \frac{1}{\lambda} = 0$$

   Hence we get

$$\lambda = \frac{\sum_{i=1}^{N} k_i}{N}$$