Generative Model and Methods and MLE

Sunday, October 28, 2018 2:09 PM

2017:

1. **[Generative Methods, MLE]**

Suppose that you are a myrmecologist (a person who studies ants), and you are currently exploring two large ant nests which are near each other in a forest.
You label the two nests '+' and '-', and let their locations be $\mu^+, \mu^- \in \mathbb{R}^2$ respectively. You observe that the positions of the ants around their own nest seem to follow spherical Gaussian distributions,

$$\mathcal{N}(\mu^+, \sigma^2 I), \qquad \mathcal{N}(\mu^-, \sigma^2 I),$$

with the same variance $\sigma 2$ for both nests because all the ants are of the same species. Let $p+$ be the probability that any given ant is from the '+' nest, and let $p- = 1 - p+$ be the probability for the '-' nest.

a. Using a video (and with help from friends who are computer-vision experts), you painstakingly tracked 1000 ants, and determined which nest they are from. At some snapshot of the video, the positions and nests of these 1000 ants are given by

$$\left(x^{(1)}, y^{(1)}\right), \dots, \left(x^{(1000)}, y^{(1000)}\right)$$

where each $x(i) \in \mathbb{R}^2$ and each $y^{(i)} \in \{+, -\}$. Let n^+, n^- be the number of ants observed from each nest respectively. Let $[\![\cdot]\!]$ be the indicator function, and let

$$\hat{p}^+ = \frac{n^+}{1000}, \qquad \hat{p}^- = \frac{n^-}{1000},$$
$$\hat{\mu}^+ = \frac{1}{n^+}\sum_i [\![y^{(i)} = +]\!] x^{(i)}, \ \mu^- = \frac{1}{n^-}\sum_i [\![y^{(i)} = -]\!] x^{(i)}$$

be the maximum likelihood estimates (MLEs) of the nest probabilities and locations. What is the correct MLE of σ^2 from the data?

Ans:

$$\sigma^2 = \frac{1}{1000} \sum_i \left\| x^{(i)} - \hat{\mu}^{y^{(i)}} \right\|^2$$

You find a new ant wandering around at position $x \in \mathbb{R}^2$. By **substituting the MLEs** above into the **log likelihood ratio**, you determine that the ant is from the '+' nest if $\alpha^\top x + \alpha_0 > 0$ for some α, α_0. What is the value of α and α_0?

Ans:

$$\alpha = \frac{1}{\sigma^2}(\mu^+ - \mu^-), \ \alpha_0 = \frac{1}{2\sigma^2}(\|\mu^-\|^2 - \|\mu^+\|^2) + \log\frac{p^+}{p^-}$$

You find a new ant wandering around at position $x \in \mathbb{R}2$. The probability that the ant is from the '+' nest may be given by a sigmoid function

$$\mathbb{P}(+|x) = \text{sigmoid}(\theta^\top x + \theta_0) = \frac{1}{1 + e^{-(\theta^\top x + \theta_0)}}.$$

This shows that our model is a special case of logistic regression. What is the value of θ and θ_0?

Ans:

$$\theta = \frac{1}{\sigma^2}(\mu^+ - \mu^-), \ \theta_0 = \frac{1}{2\sigma^2}(\|\mu^-\|^2 - \|\mu^+\|^2) + \log\frac{p^+}{p^-}$$

Differentiating Logarithms

$y = \log f(x)$	$y = \log_a f(x)$
$\frac{dy}{dx} = \frac{f'(x)}{f(x)}$	$\frac{dy}{dx} = \frac{f'(x)}{(\log a) f(x)}$

e.g. (i) $y = \log(3x+5)$, $\frac{dy}{dx} = \frac{3}{3x+5}$ $\quad (ii)$ $y = \log x^3$, $\frac{dy}{dx} = \frac{3x^2}{x^3}$

2016:

2. **Generative Methods**

a. The PDF of a Poisson distribution with the real-valued parameter $\lambda \geq 0$ is

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

where $x \geq 0$ is an integer, $x! = 1 \cdot 2 \cdot \cdots \cdot x$ and $0! = 1$. Suppose that the training data consists of independent and identically distributed samples $x(1), x(2), \dots, x(n)$. Derive the maximum likelihood estimate (MLE) of λ given this data set. (Hint: use the log likelihood.)

Ans:

The log likelihood of the data is

$$\ell(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x^{(i)}\right)\log\lambda - \sum_{i=1}^{n}\log x^{(i)}!$$

Its gradient with respect to λ is

$$\nabla\ell(\lambda) = -n + \left(\sum_{i=1}^{n} x^{(i)}\right)/\lambda.$$

Solving $\nabla\ell(\hat{\lambda}) = 0$ then gives the sample average

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}.$$

If f''(x) <=0, means the graph is concave

To check that it is the MLE, we compute the Hessian. Note that it is negative since the $x^{(i)} \geq 0$.

$$\nabla^2\ell(\lambda) = -\left(\sum_{i=1}^{n} x^{(i)}\right)/\lambda^2 \leq 0$$

b.

Consider the **expectation-maximization (EM) algorithm** for the mixture of spherical Gaussians. Write down the strategy used to initialize the means $\mu^{(1)},\dots,\mu^{(k)}$.

Ans: Use centroids derived from the k means algorithm

c.

Suppose we have a mixture of multinomial distributions whose PDF is given by

$$P(x|p,q) = \sum_{i=1}^{k} p_i P\left(x\middle|q^{(i)}\right) = \sum_{i=1}^{k} p_i q_x^{(i)}$$

where $x\in\{1,2,\dots,d\}$, $p=(p_1,\dots,p_k)\in\mathbb{R}k$ and $q^{(1)},q^{(2)},\dots,q^{(k)}\in\mathbb{R}^d$. Here, the entries of each vector $p,q^{(1)},q^{(2)}$, $\dots,q^{(k)}$ are non-negative and sum to one. Write down the formula for the soft labels $p(i|x)$ computed during the expectation step of the EM algorithm.

Ans:

$$p(i|x) = \frac{p_i P\left(x\middle|q^{(i)}\right)}{\sum_{j=1}^{k} p_j P\left(x\middle|q^{(j)}\right)} = \frac{p_i q_x^{(i)}}{\sum_{j=1}^{k} p_j q_x^{(j)}}$$

2015:

[Maximum Likelihood Estimator][MLE]

1. Consider the following distribution function (a special case of a so-called gamma distribution). It is defined for positive real numbers x > 0.

$$p(x) = cx^2\exp(-\beta x)\beta^3$$

compute the maximum likelihood estimator for parameter for given data $x_1, \dots, x_n$. We assume that the samples x_i are independent. Hint: compute the maximum likelihood estimator as the maximum of the log-likelihood, that means after applying a logarithm.

Ans: Note: product rule of log => log(M*N) = log M + log N