67750: Advanced Practical Machine Learning 2017/8

Solution 2 - Image Denoising Theoretic Solutions

Prof. Yair Weiss *TA: Daniel Gissin*

1 Theoretical Questions

Note that this section is not graded and you do not have to hand it in.

1.1 MLE Calculation

In class we calculated the MLE of a sample drawn from a Bernoulli distribution and a Gaussian distribution.

You are now the manager of the Rothberg cafeteria, and you collect N samples of the time it took from the moment person i asked for a salad until he got his salad. You assume, as one does, that the service time is distributed exponentially ($p(x) = \lambda e^{-\lambda x}$ for $x \geq 0$). Calculate the parameter λ using the MLE.

Solution

We write the likelihood function of our sample:

$$L(S;\lambda) = \prod_{i=1}^{N} \lambda e^{-\lambda x_i} = \lambda^N e^{-\lambda \sum_{i=1}^{N} x_i}$$

Now, we take the log of the likelihood function:

$$LL(S,\lambda) = Nlog(\lambda) - \lambda \sum_{i=1}^{N} x_i$$

Finally, we take the derivative w.r.t. our parameter λ and set it equal to zero:

$$\frac{\partial LL}{\partial \lambda} = \frac{N}{\lambda} - \sum_{i=1}^{N} x_i = 0 \rightarrow \hat{\lambda} = \frac{N}{\sum_{i=1}^{N} x_i} = \frac{1}{\bar{x}}$$

1.2 MLE in the EM algorithm

In class we saw that the Expectiation of the log likelihood of the Gaussian Mixture Model can be written as:

$$\mathbb{E}[LL(S,Z,\theta)] = const + \sum_{i=1}^{n} \sum_{y=1}^{k} c_{i,y} log(\pi_y N(x_i;\mu_y,\Sigma_y))$$

We then were able to maximize the above expectation w.r.t. the parameters of the GMM. Show that the MLE of the multinomial distribution of our latent variable is:

$$\pi_y = \frac{1}{n}\sum_{i=1}^{n} c_{i,y}$$

Solution

The π's are a multinomial distribution, and so they need to satisfy the following constraints:

$$\pi_y \geq 0$$

$$\sum_{y=1}^{k} \pi_y = 1$$

We'll construct the Lagrangian, adding the second constraint (we will see that our result also satisfies the first constraint):

$$L(S,\theta,\lambda) = const + \sum_{i=1}^{n}\sum_{y=1}^{k} c_{i,y}(log(\pi_y) + log(N(x_i;\mu_y,\Sigma_y)) - \lambda(\pi^T \mathbf{1} - 1)$$

We only care about the π's in this question (the other parameters are also straightforward), and so we differentiate w.r.t. π:

$$\frac{\partial L}{\partial \pi_y} = \frac{\sum_{i=1}^{n} c_{i,y}}{\pi_y} - \lambda = 0 \rightarrow \pi_y = \frac{\sum_{i=1}^{n} c_{i,y}}{\lambda}$$

We find λ using our constraint ($\sum_{y=1}^{k} \pi_y = 1$), which simply makes λ a normalizing constant, and we get:

$$\pi_y = \frac{\sum_{i=1}^{n} c_{i,y}}{\sum_{i=1}^{n}\sum_{y=1}^{k} c_{i,y}} = \frac{1}{n}\sum_{i=1}^{n} c_{i,y}$$

1.3 Multivariate Gaussian Optimal Estimator

We denote the clean signal by x, and the noisy signal by y. The relationship between them is given by:

$$y = x + \eta$$

where we assume η is Gaussian with zero mean and covariance matrix $\sigma^2 I$. For this exercise we will assume that σ is also known. As we discussed in class, this means that $y|x$ is Gaussian with mean x and covariance $\sigma^2 I$. This means that after we learn $p(x)$ we can form the joint probability $p(x,y)$ and use this to calculate the optimal estimate of x given y which we proved in class is given by $E(x|y)$.

In class we showed an analytical form for $E(x|y)$ when $p(x)$ is a scalar Gaussian:

$$x^* = \frac{\frac{1}{\sigma_x^2}\mu_x + \frac{1}{\sigma^2}y}{\frac{1}{\sigma_x^2} + \frac{1}{\sigma^2}}$$

Now let's assume that $p(x)$ is a multivariate Gaussian with mean μ and covariance Σ. Show that the optimal estimate of x given y is given by

$$x^* = \left(\Sigma^{-1} + \frac{1}{\sigma^2}I\right)^{-1}\left(\Sigma^{-1}\mu + \frac{1}{\sigma^2}y\right)$$

You may use the trick we saw in class for Gaussian distributions: $\mathbb{E}[x|y] = argmax_x(p(x|y)) = argmax_x(p(x,y))$

Solution

The pdf of $x \in \mathbb{R}^d$ is given by:

$$p(x) = \frac{1}{z_x}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

And the pdf of $y|x$ is:

$$p(y|x) = \frac{1}{z_{y|x}}e^{-\frac{1}{2\sigma^2}(x-y)^T(x-y)}$$

Where z_x and $z_{y|x}$ are normalization factors. Therefore, the joint pdf is given by:

$$p(x,y) = p(x)p(y|x) = \frac{1}{z_x z_{y|x}}e^{-\frac{1}{2\sigma^2}(x-y)^T(x-y)-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

As shown in class, in the case that x, y are Gaussian, we have that:

$$E[x|y] = \arg\max_x p(x,y) = \arg\min_x \left\{\frac{1}{2\sigma^2}(x-y)^T(x-y) + \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

Taking the derivative w.r.t. x and equating to 0 yields the desired solution:

$$\frac{\partial}{\partial x}\left\{\frac{1}{2\sigma^2}(x-y)^T(x-y) + \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} = \Sigma^{-1}(x-\mu) + \frac{1}{\sigma^2}(x-y) = 0$$

$$\implies \left(\Sigma^{-1} + \frac{1}{\sigma^2}I\right)x^* = \left(\Sigma^{-1}\mu + \frac{1}{\sigma^2}y\right) \implies x^* = \left(\Sigma^{-1} + \frac{1}{\sigma^2}I\right)^{-1}\left(\Sigma^{-1}\mu + \frac{1}{\sigma^2}y\right)$$