CSCI739 Homework 2 zl9901

1) The likelihood is a Gaussian with known variance but unknown mean, the prior belief can be expressed in terms of a gaussian distribution, such that $\mu \sim N(\mu_0, \sigma_0^2)$ where μ_0 and σ_0^2 are known.

The prior should be: $f(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}}$

The likelihood should be: $f(x \mid \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

The posterior is shown as followed:

$$f(\mu \mid x) \propto f(x \mid \mu) f(\mu) =$$

$$\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}\} * (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}$$

$$\propto \exp\{\frac{-\sigma^2\mu^2 - n\sigma_0^2\mu^2 + 2\mu(\mu_0\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i) - \mu_0^2\sigma^2 - \sigma_0^2\sum_{i=1}^{n} x_i^2}{2\sigma_0^2\sigma^2}\}$$

$$\propto \exp\{\frac{-\mu^2 + 2\mu\frac{\mu_0\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i}{\sigma^2 + n\sigma_0^2} - (\frac{\mu_0\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i}{\sigma^2 + n\sigma_0^2})^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2 + n\sigma_0^2}}\} * \exp\{\frac{\mu_0^2\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i^2}{2\sigma^2\sigma_0^2}\}$$

$$\propto \exp\{\frac{-(\mu - \frac{\mu_0\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i}{\sigma^2 + n\sigma_0^2})^2}{2 * \frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0^2}}\}$$

We let $\sigma_1^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0^2}$ and $\mu_1 = \frac{\mu_0\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i}{\sigma^2 + n\sigma_0^2}$, so that

$$f(\mu \mid x) \propto \exp\{\frac{-(\mu-\mu_1)^2}{2\sigma_1^2}\}$$

So the parameter of posterior should be as followed:

$$\mu_1 = \frac{\mu_0\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i}{\sigma^2 + n\sigma_0^2} \qquad \sigma_1^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0^2}$$

The mean of posterior is influenced by both likelihood and prior.

The variance of posterior can also be expressed as:

$$\sigma_1^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

We can conclude that when n is very large, the effect of prior can be ignored, because $\frac{n}{\sigma^2}$ can be very large.

2) Since MLE is based on likelihood,

$$f(X\mid\theta) = \prod_{i=1}^{n} f(x_i\mid\theta) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\}$$

We should take log first,

$$\log(f(X\mid\theta)) = \sum_{i=1}^{n} -\log\sqrt{2\pi\sigma^2} - \frac{(x_i-\theta)^2}{2\sigma^2}$$

Taking derivative with respect to θ, then set function value to 0

$$\frac{\partial \log(f(X\mid\theta))}{\partial\theta} = \frac{\sum_{i=1}^{n} x_i - n\theta}{\sigma^2} = 0$$

$$\theta_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Since MAP is based on the product of likelihood and prior

$$f(\theta\mid X) \propto f(X\mid\theta)f(\theta)$$

$$= \prod_{i=1}^{n} f(x_i\mid\theta)f(\theta)$$

$$= (\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(x_i-\theta)^2}{2\sigma^2}\}) * \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\{-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}\}$$

We should take log first,

$$\log(f(\theta|X)) = (\sum_{i=1}^{n} -\log(\sqrt{2\pi\sigma^2}) - \frac{(x_i-\theta)^2}{2\sigma^2}) - \log(\sqrt{2\pi\sigma_1^{\ 2}})$$

Taking derivative with respect to θ, then set function value to 0

$$\frac{\partial \log(f(\theta|X))}{\partial\theta} = \frac{\sum_{i=1}^{n} x_i - n\theta}{\sigma^2} - \frac{\theta-\mu_1}{\sigma_1^2} = 0$$

$$\theta_{MAP} = \frac{\sigma^2\mu_1 + \sigma_1^2\sum_{i=1}^{n} x_i}{\sigma^2 + n\sigma_1^2}$$

3) Prior: $\mu_0 = 70, \sigma_0 = 5$

Likelihood: $\mu = 75, \sigma = 10$

$$\theta_{MAP} = \frac{\sigma^2\mu_0 + \sigma_0^2\sum_{i=1}^{n} x_i}{\sigma^2 + n\sigma_0^2} = \frac{10^2 * 70 + 5^2 * 75 * 30}{10^2 + 30 * 5^2} \approx 74.41$$

Because θ_{MAP} is closer to the mean of likelihood which is 75. So likelihood has much effect than prior.

In addition, if likelihood is not included, we can just use prior to make decisions. After we introduce the concept of likelihood and according to the knowledge of Bayesian, we can get the modified result which is also posterior. So likelihood really matters. We can use the conclusion of question a to prove that likelihood has greater influence than prior.

4) Please see the code attached to this zip file.