COGS 118A: ASSIGNMENT 4 SOLUTIONS

Problem 1. Maximum likelihood estimation for Gaussians (40)

(a) (5) For $x_1 \ldots, x_n \sim N(\mu^*, (\sigma^*)^2)$

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

$$\mathbb{E}[\mu] = \mathbb{E}[\frac{1}{n} \sum_{i=1}^{n} x_i]$$

$$\mathbb{E}[\mu] = \frac{1}{n} \mathbb{E}[\sum_{i=1}^{n} x_i]$$

$$\mathbb{E}[\mu] = \frac{1}{n} \sum_{i=1}^{n} \mu^*$$

$$\mathbb{E}[\mu] = \frac{1}{n} \cdot n \cdot \mu^* = \mu^*$$

(b) (5) We can find the maximum likelihood estimate for the variance of the Gaussian distribution that generated the data in (a). It is the sample variance (proof omitted):

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Show that the MLE estimate for the variance has the following expectation:

$$\mathbb{E}[\Sigma^2] = \frac{n-1}{n} (\sigma^2)^*$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \sum_{i=1}^{n} [(x_i - \mu + \mu - \mu^*)^2]$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \sum_{i=1}^{n} [(x_i - \mu)^2 + 2(x_i - \mu)(\mu - \mu^*) + (\mu - \mu^*)^2]$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \sum_{i=1}^{n} [(x_i - \mu)^2 + 2x_i \mu - 2x_i \mu^* \mu \mu^* + (\mu - \mu^*)^2]$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} [(x_i - \mu)^2] = \frac{1}{n} \sum_{i=1}^{n} [(x_i - \mu^*)^2] - n(\mu - \mu^*)^2$$

$$\mathbb{E}[\Sigma^2] = \mathbb{E}[\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu^*)^2]$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} (x_i - \mu^*)^2$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} (x_i - \mu^*)^2 - n(\mu - \mu^*)^2$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \mathbb{E} [x_i - \mu^*] - \mathbb{E} [\mu - \mu^*]^2$$

$$\mathbb{E}[\Sigma^2] = \frac{1}{n} \cdot n \sigma^2 - \text{Var}(\sum_{i=1}^{n} X_i)$$

$$\mathbb{E}[\Sigma^2] = \sigma^2 - \frac{\sigma^2}{n}$$

$$\mathbb{E}[\Sigma^2] = \frac{n-1}{n} \sigma^2$$
(c) (5) We define a new estimator for the variance

\[ \tilde{\sigma}^2 = \frac{n}{n-1} \Sigma^2 \]
\[ \mathbb{E}[\tilde{\sigma}^2] = \sigma^2 \]
\[ \mathbb{E}[\tilde{\sigma}^2] = \mathbb{E}\left[ \frac{n}{n-1} \Sigma^2 \right] \]
\[ \mathbb{E}[\tilde{\sigma}^2] = \frac{n}{n-1} \mathbb{E}[\Sigma^2] \]
\[ \mathbb{E}[\tilde{\sigma}^2] = \frac{n}{n-1} \cdot \frac{n-1}{n} \sigma^*^2 = \sigma^*^2 \]

(d) (10) (Matlab) (Figure 1a)

(e) (10) (Matlab) (Figure 1b)

Our variance over data sets is higher because we are using less data sets and overall the data is less average.

(f) (5) (Matlab) (Figure 2) As we increase the number of data points used to calculate the mean we see the average mean converge to the expected value zero. As expected per central limit theory.

**Problem 2. Bayesian inference for Gaussians (35)**

(a) (10) For an iid set with data points \( x_1, \ldots, x_n \sim \mathcal{N}(\mu, \sigma^2) \), where \( \sigma^2 \) is known, and \( \mu \) is unknown. Prior Gaussian distribution \( p(\mu) = \mathcal{N}(\mu_0, \sigma_0^2) \). Use Bayes’ Rule to show that the posterior distribution is also Gaussian, \( p(\mu|x_1, \ldots, x_n) = \mathcal{N}(\mu_n, \sigma_n^2) \), i.e. the dependence of the posterior on \( \mu \) is quadratic in the exponent. By completing the square, show that the precision of the posterior is the sum of the precision of the prior and likelihood
Figure 2: sequentially estimated $\mu_{MLE}$

- Prior: $p(\mu) = \mathcal{N}(\mu_0, \sigma^2)$
  
  $$p(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\mu - \mu_0)^2\right]$$

- Likelihood: $\prod^n p(x_i|\mu, \sigma^2) = \prod^n \mathcal{N}(x_i; \mu, \sigma^2)$
  
  NOTE: Gaussian distribution is $\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

- Posterior:
  
  $$p(\mu|x_1, \ldots, x_n) = \frac{p(x_1 \ldots x_n|\mu)p(\mu)}{p(x_1, \ldots, x_n)}$$
  
  $$\propto p(x_1, \ldots, x_n)p(\mu)$$
  
  $$\propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\sum^n (x_i - \mu)^2\right]$$
  
  $$\propto \exp\left[-\frac{1}{2\sigma^2}\sum^n \left(\frac{1}{2\sigma^2}(x_i - \mu)^2 - \frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{\sigma_0^2}\right)\right]$$
  
  $$= \exp\left[-\frac{n\mu^2}{2\sigma^2} + 2\frac{1}{2\sigma^2}\sum^n x_i\frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{\sigma_0^2}\right]$$
  
  $$\propto \exp\left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2}\mu^2 + \frac{1}{2\sigma^2}\sum^n x_i\frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{\sigma_0^2}\right]$$

Note: Completing the square example, We want to turn $ax^2 + 2bx$ into $c(x+y)^2 + d$

$$= a(x^2 + 2\frac{b}{a}x)$$

$$= -\left[(x + \frac{b}{a})^2 - \frac{b^2}{a^2}\right]$$

$$= a(x + \frac{b}{a})^2 - \frac{b^2}{a}$$

$$c = a; y = \frac{b}{a}; d = \frac{b^2}{a}$$

$$\propto \exp\left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2}\mu^2 + \frac{2}{2\sigma^2} \sum^n x_i + \frac{\mu_0^2}{\sigma_0^2}\right]$$
\[ \propto \exp \left[ -\frac{2\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} \right] \]

\[ \propto \exp \left[ -\frac{2\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} \right] \]

\[ \exp \left[ -\frac{2\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} \right] \]

\[ \exp \left[ -\frac{\mu - x_i \sigma}{\sigma^2 + n \sigma^2} \right] \]

\[ \mu_n = \frac{\sigma_0^2 \sum x_i + \mu \sigma^2}{\sigma^2 + n \sigma_0^2} \mu_{MLE} = \frac{1}{n} \sum x_i \]

\[ = \frac{\sigma_0^2 \mu_{MLE} + \mu_0}{\sigma^2 + n \sigma_0^2} \]

\[ = \frac{\sigma_0^2}{\sigma^2 + n \sigma_0^2} \mu + \frac{n \sigma^2}{\sigma^2 + n \sigma_0^2} \mu_{MLE} \]

\[ = \frac{1}{\sigma_0^2 + \frac{n}{\sigma^2}} \mu_0 + \frac{n}{\sigma_0^2 + \sigma^2} \mu_{MLE} \]

consequently

\[ \sigma_n^2 = \frac{\sigma_0^2}{\sigma^2 + n \sigma_0^2} \sigma^2 \]

\[ \frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \]

(b) (5) (Matlab) (Figure 3a)

(a) Problem 2b

(b) Problem 2c

Figure 3: 30 points and 1000 points

(c) (5) (Matlab) (Figure 3b)

(d) (10) (Matlab) (Figure 4)

(e) Bishop problem 2.45

(Bishop 2.155) \[ W(\Lambda|W, \nu) = B|\Lambda|^{(\nu-D-1)/2} \exp \left( -\frac{1}{2} Tr(W^{-1} \Lambda) \right) \]
Prior: \( W(\Lambda | W, D) = B|\Lambda|^{\frac{D-D-1}{2}} \exp(-\frac{1}{2} \text{Tr}(W^{-1} \Lambda)) \)

Likelihood: \( \mathcal{N}(X | \mu^{-1}) = \frac{1}{(2\pi)^{D/2}} |\Lambda|^{\frac{1}{2}} \exp(-\frac{1}{2} (X - \mu)^T \Lambda (X - \mu)) \)

Note: Posterior \( \propto \) prior \( \cdot \) Likelihood

Posterior: \( B \frac{1}{(2\pi)^{D/2}} |\Lambda|^{\frac{D-D-1}{2} + \frac{D}{2}} \exp(-\frac{1}{2} \text{Tr}(W^{-1} + (X - \mu)(X - \mu))) \)

**Problem 3.** Exponential family of distributions (25)

(a) (5) The logistic sigmoidal function is defined as
\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]
show that \( 1 - \sigma(x) = \sigma(-x) \)
\[
1 - \sigma(x) = \frac{1}{1 + e^{-x}} - 1
= \frac{1}{1 + e^{-x} - 1}
= \frac{e^{-x}}{1 + e^{-x} - 1}
= \frac{e^{-x}}{1 + e^{-x}}
= e^{-x}
\]

(b) (10) Bishop 2.57
Standard form: \( p(p | \eta) = h(x) g(\eta) exp(\eta^T u(x)) \)
Gaussian: \( \mathcal{N}(\mu, \Sigma^2) = \frac{1}{(2\pi)^{D/2}} \frac{1}{\Sigma^{1/2}} \exp\left(\frac{1}{2} (\bar{x} - \bar{\mu})^T \Sigma^{-1} (\bar{x} - \bar{\mu})\right) \)
\( \eta^T u(x) = \frac{1}{2} (\bar{x} - \bar{\mu})^T \Sigma^{-1} (\bar{x} - \bar{\mu}) \)
\[ u(x) = \begin{pmatrix} \frac{1}{2} X^T X \\ X^T \end{pmatrix} \]
\[ \eta^T = \left( \Sigma^{-1} \Sigma^{-1} \mu \right) \]

\[ h(x)g(\eta) = \frac{1}{\sqrt{2\pi} |\Sigma|} \exp\left( \frac{1}{2} \mu^T \Sigma^{-1} \mu \right) \]

\[ h(x) = \frac{1}{\sqrt{2\pi} \sigma} \cdot g(\eta) = |\Sigma|^{-1/2} \exp\left( \frac{1}{2} \mu^T \Sigma^{-1} \mu \right) \]

(c) (10) Bishop 2.58

\[ -\nabla \nabla \ln g(\eta) = \mathbb{E}[u(X)u(X)^T] - \mathbb{E}[u(X)]\mathbb{E}[u(X)^T] = \text{Cov}[u(X)] \]

\[ \nabla g(\eta) \int h(x) \exp \eta^T u(X) dx + g(\eta) \int h(x) \exp \eta^T u(X) u(X) dx = 0 \]

Note: \( g(\eta) \int h(x) \exp \eta^T u(X) dx = 1 \)

\[ \frac{1}{g(\eta)} \nabla g(\eta) + \int h(x) \exp \eta^T u(X) dx = 0 \]

\[ -\frac{1}{g(\eta)} \nabla g(\eta) = \mathbb{E}[u(X)] \]

\[ -\nabla \nabla \ln[g(\eta)] = g(\eta) \int h(x) \exp \eta^T u(X) u(X) dx \]

\[ -\nabla \nabla \ln[g(\eta)] = -\nabla g(\eta) \int h(x) \exp \eta^T u(X) u(X) dx + g(\eta) \int h(x) \exp \eta^T u(X) u(X)^T u(X) dx \]

\[ = -g(\eta) \int h(x) \exp \eta^T u(X) u(X) dx \cdot g(\eta) \int h(x) \exp \eta^T u(X) u(X) dx + g(\eta) \int h(x) \exp \eta^T u(X) u(X)^T u(X) dx \]

\[ = -\mathbb{E}[u(X)^T] \mathbb{E}[u(X)] - \mathbb{E}[u(X) u(X)^T] \]

\[ -\nabla \nabla \ln[g(\eta)] = \text{Cov}[u(X)] \]