Problem Set 2
ECE 685D Fall 2020

Name: Guillem Amat Castello

Problem 1: Conditional Multivariate Gaussian Distribution

All conditional distributions of a multivariate normal distribution are normal. Therefore, in order to derive the conditional pdf of x_2 given $x_1 = a$, we will need to find the variance and the mean vector.

We will start by defining a linear combination $z = x_1 + Ax_2$, where the vector z is independent with respect to x_1 and where $A = -\Sigma_{x_2x_1}\Sigma^{'}_{x_2x_2}$.

$$\begin{aligned}
\mu_{x_2|x_1} &= E(z - Ax_1|x_2) \\
&= E(z) - Ax_1 \\
&= \mu_{x2} + A\mu_1 - Ax_1 \\
&= \mu_{x2} + A(\mu_1 - x_1) \\
&= \mu_{x2} + \Sigma_{x_2x_1}\Sigma^{'}_{x_2x_2}(x_1 - \mu_1)
\end{aligned} \tag{1}$$

$$\begin{aligned}
\Sigma_{x_2|x_1} &= \hat{V}(z) \\
&= \hat{V}(x_2) + A\hat{V}(x_1)A^{'} + A\sigma(x_2, x_1) + \sigma(x_1, x_2)A^{'} \\
&= \Sigma_{x_2x_2} + \Sigma_{x_2x_1}\Sigma^{'}_{x_2x_2}\Sigma_{x_1x_2} - 2\Sigma_{x_2x_1}\Sigma^{'}_{x_2x_2}\Sigma_{x_1x_2} \\
&= \Sigma_{x_2x_2} - \Sigma_{x_2x_1}\Sigma^{'}_{x_2x_2}\Sigma_{x_1x_2}
\end{aligned} \tag{2}$$

When $x_1 = a$, then $\mu_{x_2|x_1} = \mu_{x_2}$ and $\Sigma_{x_2|x_1} = \Sigma_{x_2x_2}$. The pdf will be given by:

$$f(x_2|x_1) = (2\pi)^{-D/2}det(\Sigma_{x_2x_2})^{-0.5}e^{-0.5(x-\mu_2)T\Sigma^{'}_{x_2x_2}(x-\mu_2)}$$

We can calculate the mean vector substituting in (1):

$$\mu_{x_2|x_1=a} = \mu_{x2} - \Sigma_{x_2x_1}\Sigma^{'}_{x_2x_2}(\mu_1 - x_1) = \begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}0.25 & 0\\0 & 0.25\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix}1\\2\end{bmatrix}$$

We can calculate the covariance matrix substituting in (2):

$$\Sigma_{x_2|x_1=a} = \Sigma_{x_2x_2} - \Sigma_{x_2x_1}\Sigma^{'}_{x_2x_2}\Sigma_{x_1x_2} = \begin{bmatrix}1.25 & 0\\0 & 1.25\end{bmatrix} - \begin{bmatrix}0.25 & 0\\0 & 0.25\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}0.25 & 0\\0 & 0.25\end{bmatrix}$$
$$= \begin{bmatrix}1.1875 & 0\\0 & 1.1875\end{bmatrix}$$

Problem 3: Bayesian linear regression

We can use a Multivariate Normal Bayesian Linear Regression to estimate the location of the sensor measurements. We will start by stating Bayes rule and then developing it in the context of this problem:

$$\pi(\beta \mid y, x, \sigma_2) = p(y \mid x, \beta, \sigma_2)\ \pi(\beta)$$

The posterior predictive distribution would be given by the following expression:

$$\pi(\beta \mid y, x, \sigma_2) = \int \Big[p(y \mid x, \beta, \sigma_2) d\ \pi(\beta) \Big]\ d\theta = \mathbb{E}_\theta[p(y|x,\theta)]$$

Which we can interpret as the average prediction for all plausible parameters θ according to the prior distribution.

The likelihood will be given by $y_i \sim \mathcal{N}(x, \Sigma_n)$, we develop it below:

$$\begin{aligned} p(y \mid x, \beta, \sigma_2) &\propto (\sigma^2)^{\frac{-n}{2}} exp\Big\{ -\frac{1}{2\sigma^2}[y^T y - 2\beta^T x^T y + \beta^T x^T x \beta] \Big\} \\ &\propto exp\Big\{ -\frac{1}{2\sigma^2}[\beta^T x^T x \beta - 2\beta^T x^T x y] \Big\} \\ &\propto exp\Big\{ -\frac{1}{2}\beta^T \Big(\frac{1}{\sigma^2} x^T x \Big)\beta + \beta^T \Big(\frac{1}{\sigma^2} x^T y \Big) \Big\} \end{aligned} \tag{3}$$

Our prior will be given by $x \sim \mathcal{N}(\mu_0, \Sigma_0)$, we develop it below:

$$\pi(\beta) \propto exp\Big\{ -\frac{1}{2}\beta^T \Sigma_0^{'} \beta + \beta^T \Sigma_0^{'} \mu_0 \Big\} \tag{4}$$

We are now in a position to derive the posterior distribution in the following way:

$$\begin{aligned} \pi(\beta \mid y, x, \sigma_2) &\propto p(y \mid x, \beta, \sigma_2)\ \pi(\beta) \\ &\propto exp\Big\{ -\frac{1}{2}\beta^T \Big[\Sigma_0^{'} + \frac{1}{\sigma^2} x^T x \Big] \beta + \beta^T \Big[\Sigma_0^{'} \beta_0 + \frac{1}{\sigma^2} x^T y \Big] \Big\} \\ &\equiv \mathcal{N}(\mu_n, \Sigma_n) \end{aligned} \tag{5}$$

We find that a Gaussian prior leads to a Gaussian posterior. The Gaussian Distribution is the conjugate prior of Linear Regression. Comparing this to the prior we get the covariance matrix and the mean:

$$\Sigma_n = \Big[\Sigma_0^{'} + \frac{1}{\sigma^2} x^T x \Big]^{'}$$

$$\mu_n = \Sigma_n \Big[\Sigma_0^{'} \beta_0 + \frac{1}{\sigma^2} X^T y \Big]$$

To estimate the location of the sensor I would use MLE:

$$\begin{aligned}
\pi(\beta \mid y, x, \sigma_2) &= p(y \mid x, \beta, \sigma_2)\ \pi(\beta) \\
&= \prod_{i=1}^{N} p(y_i \mid x_i, \beta, \sigma_2) \pi(\beta) \\
&= \prod_{i=1}^{N} \mathcal{N}(\mu_i, \Sigma_i)
\end{aligned} \tag{6}$$

Problem 5: Minimizing Minkowski Loss

1. General E[L] minimization:

In order to minimize the expected value of the loss function, we will need to compute the derivative of $\mathrm{E}[L]$ with respect to $y(x)$.

We know that our loss function is $L(t, y(x)) = |t - y(x)|$, that $p(x,t) = p(t|x)p(x)$ given the law of total probability and that $p(x)$ is non-negative and independent with regards to $y(x)$. With this, we can derive:

$$\begin{aligned}
\mathrm{E}[L] &= \iint \Big[L(t, y(x))\, p(x,t)dt\Big]dx \\
&= \iint \Big[|t - y(x)|\, p(t|x)dt\Big]p(x)dx \\
&= \int_{y_x}^{\infty} \Big[|t - y(x)|\, p(t|x)dt\Big] + \int_{-\infty}^{y_x} \Big[|t - y(x)|\, p(t|x)dt\Big]
\end{aligned} \tag{7}$$

In (7), $p(x|t)t$ does not depend on $y(x)$, therefore we can disregard it. We will continue by differentiating the expression above with respect to $y(x)$ and setting it to zero:

$$\frac{\partial \mathrm{E}[L]}{\partial y(x)} = \int_{y_x}^{\infty} p(t|x)dt + \int_{-\infty}^{y_x} p(t|x)dt = 0 \tag{8}$$

The value that minimizes the expression above is the median.

2. Provided E[L] minimization:

Given the loss function provided in the problem statement, we can substitute in (8) to get:

$$\frac{\partial \mathrm{E}[L]}{\partial y(x)} = \int_{y_x - \delta}^{\infty} p(t|x)dt + \int_{-\infty}^{\delta - y_x} p(t|x)dt = 1 - \int_{y_x - \delta}^{y_x + \delta} p(t|x)dt$$