CSCI3320 Homework 4 1155110208 / Chung Tsz Ting

1) As the probability density function of multivariate normal distribution is $p(x) = \frac{1}{2\pi^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma(x-\mu)\right]$ for $\mu^T = [\mu_1, \mu_2]$ and $\Sigma = \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$.

$$
\begin{aligned}
&p(x) \\
&= \frac{1}{2\pi^{\frac{2}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \\
&= \frac{1}{2\pi\cdot\det\left(\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\right)^{1/2}}\exp\left[-\frac{1}{2}[x_1-\mu_1 \quad x_2-\mu_2]\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}^{-1}\begin{bmatrix}x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}\right] \\
&= \frac{1}{2\pi(\sigma_1^2\sigma_2^2-\rho^2\sigma_1^2\sigma_2^2)^{1/2}}\exp\left[-\frac{1}{2}[z_1\sigma_1 \quad z_2\sigma_2]\frac{1}{\det\Sigma}\begin{bmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{bmatrix}\begin{bmatrix}z_1\sigma_1 \\ z_2\sigma_2\end{bmatrix}\right] \\
&= \frac{1}{2\pi((1-\rho^2)\sigma_1^2\sigma_2^2)^{1/2}}\exp\left[-\frac{1}{2det\Sigma}\,[z_1\sigma_1\sigma_2^2-\rho z_2\sigma_1\sigma_2^2 \quad z_2\sigma_2\sigma_1^2-\rho z_1\sigma_2\sigma_1^2]\begin{bmatrix}z_1\sigma_1 \\ z_2\sigma_2\end{bmatrix}\right] \\
&= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left[-\frac{1}{2(1-\rho^2)\sigma_1^2\sigma_2^2}(z_1^2\sigma_1^2\sigma_2^2-\rho z_1z_2\sigma_1^2\sigma_2^2+z_2^2\sigma_2^2\sigma_1^2-\rho z_1z_2\sigma_1^2\sigma_2^2)\right] \\
&= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left[-\frac{1}{2(1-\rho^2)^{\frac{1}{2}}}(z_1^2-\rho z_1z_2+z_2^2-\rho z_1z_2)\right] \\
&= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left[-\frac{1}{2(1-\rho^2)^{\frac{1}{2}}}(z_1^2-2\rho z_1z_2+z_2^2)\right]
\end{aligned}
$$

2) To make a quadratic fit using the two variable x_1 and x_2, we first have the model of linear regression $g(x|\theta) = w_0 + w_1y_1 + w_2y_2 + w_3y_3 + w_4y_4 + w_5y_5$where $y_3 = x_1,\ y_2 = x_2,$ $y_3 = x_1x_2,\ y_4 = (x_1)^2$ and $y_5 = (x_2)^2$ with parameters w_i from i=0 to 5 be randomly initialized. Then with the given sample of $\chi = \{x_1^t, x_2^t, r^t\}$, the parameters can be optimized by calculating the expected outcome by $g(x|\theta)$ and performing gradient descent to find the minimum of the loss function i.e. $\mathcal{L}(\theta) = \frac{1}{2t}\Sigma(g(x^t|\theta) - r^t)^2$. [where the optimization equation is $w_i = w_i - \alpha\frac{d\mathcal{L}(\theta)}{dw_i}$]