Assignment 1 Solutions

1 SVD and Eigendecomposition

1.1 SVD and Eigendecomposition Basics

a) $A^\top = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$

$A^\top A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

b) First method: We could just do the eigendecomposition of $A^\top A$, and show that the eigenvalues are the same, and the eigenvectors are the same up to a multiplicative constant.

Second method:

- Check that U is an orthogonal matrix by showing inner products between different columns are 0.
- Check that $A^\top A$ times any column of U gives us the same vector scaled by the corresponding eigenvalue.
- Check that the first matrix is the inverse of the last (and that the middle matrix is diagonal).
- Check that the product of the three matrices equals to $A^\top A$.

c) As the singular values σ_i are the square roots of eigenvalues of $A^\top A$, which are 6 and 1 we can get from b). Since rank(A) = 2, there are only two nonzero singular values: $\sigma_1 = \sqrt{6}$ and $\sigma_2 = \sqrt{1}$. The singular value matrix Σ must be the same size as A, so we have

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{1}$$

1.2 Geometric Interpretation of SVD

a) $U = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$ can be written as $U = \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix}$.

Therefore, $\theta_U = \frac{\pi}{4}$, U is a rotation matrix.

$V^\top = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^\top$ can be written as $V^\top = \begin{bmatrix} \cos -\frac{\pi}{6} & -\sin -\frac{\pi}{6} \\ \sin -\frac{\pi}{6} & \cos -\frac{\pi}{6} \end{bmatrix}$.

Therefore, $\theta_V^\top = -\frac{\pi}{6}$, $V^\top$ is a rotation matrix.

b) First, $V^\top$ performs a clockwise rotation by $\frac{\pi}{6}$. Next, the singular matrix performs scaling by a factor of 2 in the x direction, and by a factor of $\frac{1}{2}$ in the y direction. Finally, U performs a counter-clockwise rotation by $\frac{\pi}{4}$.

2 Convexity and Linear Algebra

2.1 Taylor Expansions

a) Computing the gradient and Hessian is an exercise in taking partial derivatives.

$$\frac{\partial f}{\partial \vec{x}} = \begin{pmatrix} 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 2x_3 \end{pmatrix} \tag{2}$$

$$\frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^\top} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix} \tag{3}$$

b) The second order Taylor expansion is given by

$$g(\vec{x}) = \vec{x}_0 + \frac{\partial f}{\partial \vec{x}}(\vec{x}_0)(\vec{x} - \vec{x}_0) + \vec{x}^\top \frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^\top}(\vec{x}_0)\vec{x} \tag{4}$$

In this case, $\frac{\partial f}{\partial \vec{x}}(\vec{x}_0) = \vec{0}$, and $\frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^\top}(\vec{x}_0) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}$, so the Taylor expansion is

$$g(\vec{x}) = \frac{1}{2}\left(4x_1^2 + 2x_2^2 + 4x_2x_3 + 2x_3^2\right) \tag{5}$$

c) The Hessian matrix $\frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^\top}$ is positive semi-definite everywhere. Hence, f is convex.

2.2 Matrix Rank and Inverse

a) A is full rank $\Rightarrow$ its columns are linearly independent.

Thus, if $\vec{x} = \vec{0}$ then $A\vec{x} = \vec{0}$, and if $A\vec{x} = \vec{0}$ then $\vec{x} = \vec{0}$, since $A\vec{x}$ is a linear combination of the columns of A.

b)

$$\vec{x}^\top (A^\top A)\vec{x} = (A\vec{x})^\top (A\vec{x}) \tag{6}$$

$$> 0 \text{ if } \vec{x} \neq \vec{0} \text{ since we have already shown } A\vec{x} = \vec{0} \Leftrightarrow \vec{x} = \vec{0} \tag{7}$$

Therefore, $A^\top A$ is positive definite.

c) A symmetric matrix has an eigendecomposition of $U\Lambda U^\top$, where U is orthogonal and Λ is diagonal.

If the matrix is positive definite, then all diagonal elements of Λ are positive.

Therefore, the inverse exists and is given by $U\Lambda^{-1}U^\top$, where Λ^{-1} is constructed by taking the reciprocal of the diagonal elements of Λ.

2.3 The Normal Equations

a) In general, we have $\vec{x} = (A^\top A)^{-1} A^\top \vec{b}$.

When $m = n$ and A is full rank, then A is invertible, and we have $(A^\top A)^{-1} = A^{-1}(A^\top)^{-1}$, so

$$\begin{aligned}
\vec{x} &= (A^\top A)^{-1} A^\top \vec{b} && \text{(8a)}\\
&= A^{-1}(A^\top)^{-1} A^\top \vec{b} && \text{(8b)}\\
&= A^{-1}\vec{b} && \text{(8c)}
\end{aligned}$$

b) $A^\top \vec{b}$ represents a linear combination of the columns of $A^\top$, and $A^\top A\vec{x}$ represents a linear combination of the columns of $A^\top A$.

Therefore, we would like to show that any linear combination of the columns of $A^\top$ can be written as a linear combination of the columns of $A^\top A$. This is equivalent to showing $R(A^\top) \subseteq R(A^\top A)$.

To show this, we write $\vec{b} = \vec{b}_1 + \vec{b}_2$, where $\vec{b}_1 \in R(A)$ and $\vec{b}_2 \in N(A^\top)$. Then,

$$\begin{aligned}
A^\top \vec{b} &= A^\top \vec{b}_1 + \vec{b}_2 && \text{(9a)}\\
&= A^\top \vec{b}_1 \text{ since } \vec{b}_2 \in N(A^\top) && \text{(9b)}\\
&= A^\top A\vec{x} \text{ for some } \vec{x},\ \text{since } b_1 \in R(A) && \text{(9c)}
\end{aligned}$$

Therefore $R(A^\top) \subseteq R(A^\top A)$ and there always exists a solution to $A^\top A\vec{x} = A^\top \vec{b}$. In fact, $R(A^\top) = R(A^\top A)$ since $R(A^\top A) \subseteq R(A^\top)$ always.

Another way to prove this, using different linear algebra facts, is to use the reduced SVD. Let r be the rank of A, then the reduced SVD of A is given by $A = U_r\Sigma_r V_r^\top$, where $U_r \in \mathbb{R}^{m\times r}, \Sigma_r \in \mathbb{R}^{r\times r}, V_r \in \mathbb{R}^{r\times n}$ are all full rank. The normal equation now becomes

$$\begin{aligned}
A^\top A\vec{x} &= A^\top \vec{b} && \text{(10a)}\\
(U_r\Sigma_r V_r^\top)^\top (U_r\Sigma_r V_r^\top)\vec{x} &= (U_r\Sigma_r V_r^\top)^\top \vec{b} && \text{(10b)}\\
V_r\Sigma_r^\top U_r^\top U_r\Sigma_r V_r^\top \vec{x} &= V_r\Sigma_r^\top U_r^\top \vec{b} && \text{(10c)}\\
V_r\Sigma_r I_r\Sigma_r V_r\vec{x} &= V_r\Sigma_r U_r^\top \vec{b}, \quad \text{since } U_r^\top U_r = I_r\ (r\times r \text{ identity matrix}) && \text{(10d)}\\
V_r\Sigma_r\Sigma_r V_r\vec{x} &= V_r\Sigma_r U_r^\top \vec{b} && \text{(10e)}\\
V_r^\top V_r\Sigma_r\Sigma_r V_r\vec{x} &= V_r^\top V_r\Sigma_r U_r^\top \vec{b} && \text{(10f)}\\
I_r\Sigma_r\Sigma_r V_r\vec{x} &= I_r\Sigma_r U_r^\top \vec{b}, \quad \text{since } V_r^\top V_r = I_r && \text{(10g)}\\
\Sigma_r^{-1}\Sigma_r\Sigma_r V_r\vec{x} &= \Sigma_r^{-1}\Sigma_r U_r^\top \vec{b} && \text{(10h)}\\
\Sigma_r V_r\vec{x} &= U_r^\top \vec{b} && \text{(10i)}
\end{aligned}$$

Let $\tilde{A} = \Sigma_r V_r$, $\tilde{\vec{b}} = U_r^\top \vec{b}$, then Eq. (10i) becomes the system of equations $\tilde{A}\vec{x} = \tilde{\vec{b}}$ where $\tilde{A}$ is full rank. Since $\tilde{A}$ is full rank, its rank cannot be increased by adding a column, so the rank of $\begin{bmatrix} \tilde{A} & \tilde{\vec{b}} \end{bmatrix}$ is the same as the rank of $\tilde{A}$. This implies that there exists a solution $\vec{x}$ that satisfies $\tilde{A}\vec{x} = \tilde{\vec{b}}$, or equivalently, $A^\top A\vec{x} = A^\top \vec{b}$. The last fact can be shown by construction via row reducing $\begin{bmatrix} \tilde{A} & \tilde{\vec{b}} \end{bmatrix}$.

3 Probability

3.1 Conditional Bayes' Rule

a)

By the Definition of Conditional Probability

$$p(x, y, z) = p(x, y|z)p(z) \Leftrightarrow \frac{p(x, y, z)}{p(z)} = p(x, y|z) \tag{11a}$$

By the General Product Rule

$$p(x, y, z) = p(x|y, z)p(y, z) = p(x|y, z)p(y|z)p(z) \tag{11b}$$

Combining (10a) and (10b)

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(x|y, z)p(y|z)p(z)}{p(z)} = p(x|y, z)p(y|z) \tag{11c}$$

b)

Substituting (10c) to the right hand side of the equation

$$\frac{p(x|y, z)p(y|z)}{p(x|z)} = \frac{p(x, y|z)}{p(x|z)} \tag{12a}$$

Since $p(x, y, z) = p(y, x, z)$, following the same steps in part a)

$$p(y, x|z) = p(y|x, z)p(x|z) \Leftrightarrow p(x|z) = \frac{p(y, x|z)}{p(y|x, z)} \tag{12b}$$

Noting

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(y, x, z)}{p(z)} = p(y, x|z) \tag{12c}$$

We have

$$p(x|z) = \frac{p(x,y|z)}{p(y|x,z)} \tag{12d}$$

Substituting (11d) into (11a)

$$\frac{p(x,y|z)}{p(x|z)} = \frac{p(x,y|z)p(y|x,z)}{p(x,y|z)} = p(y|x,z) \tag{12e}$$

3.2 Gaussian Distribution

a) Since the marginal distribution of a multivariate normal distribution is distributed normally, x_1 and x_2 are also distributed normally. Thus, we have

$$\begin{aligned}
p(\vec{x};\mu,\Sigma) &= \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^\top\Sigma^{-1}(\vec{x}-\vec{\mu})\right) && (13a)\\
&= \frac{1}{2\pi(\sigma_1^2\sigma_2^2)^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\begin{bmatrix}x_1-\mu_1\\x_2-\mu_2\end{bmatrix}^\top\begin{bmatrix}\sigma_1^2 & 0\\0 & \sigma_2^2\end{bmatrix}^{-1}\begin{bmatrix}x_1-\mu_1\\x_2-\mu_2\end{bmatrix}\right) && (13b)\\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2\right) && (13c)\\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2\right)\exp\left(-\frac{1}{2\sigma_2^2}(x_1-\mu_1)^2\right) && (13d)\\
&= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2\right)\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{1}{2\sigma_2^2}(x_1-\mu_1)^2\right) && (13e)\\
&= p(x_1;\mu_1,\sigma_1^2)p(x_2;\mu_2,\sigma_2^2) && (13f)
\end{aligned}$$

where the fact that x_1 and x_2 are distributed normally is used from Eq. (13e) to (13f).

b) The marginal probability of x_2 is computed by summing over x_1. Since we have $\vec{\mu}$ and Σ we can immediately determine:

$$x_2 \sim \mathcal{N}\left(\frac{1}{2}, 2\right)$$

Using the properties for conditionals of Gaussians, we have

$$x_1|x_2 = 3 \sim \mathcal{N}\left(1+\frac{1}{2}\cdot\frac{1}{2}(3-\frac{1}{2}), 1-\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right) = \mathcal{N}\left(\frac{13}{8},\frac{7}{8}\right)$$

$\mathbb{E}[x_2] = \frac{1}{2}$, $\mathbb{E}[x_1|x_2 = 3] = \frac{13}{8}$. Plotting the two distributions, we find that the conditional distribution increases the mean after observing $x_2 = 3$, the density shifts.

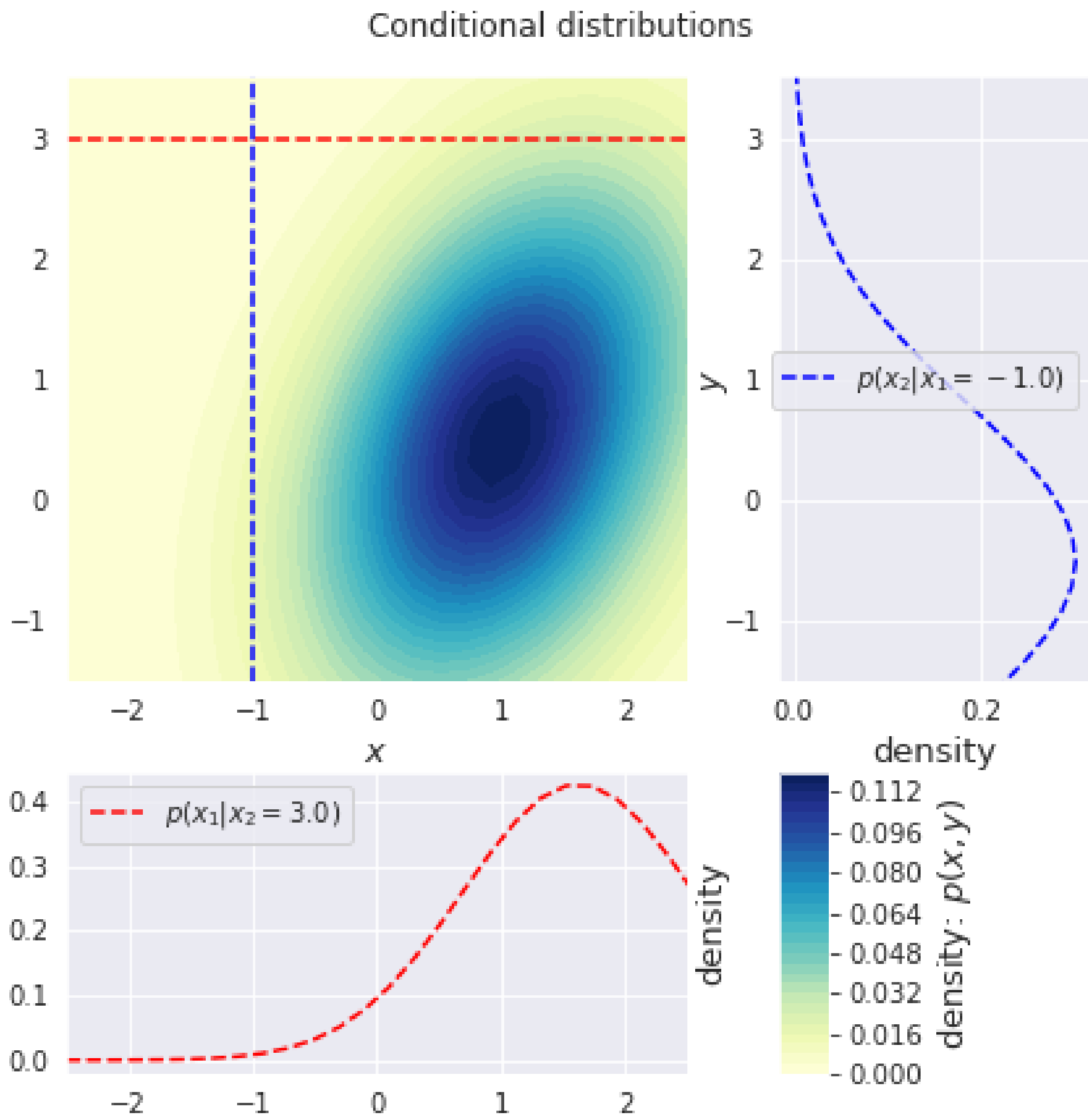

Figure 1: Conditional Distribution

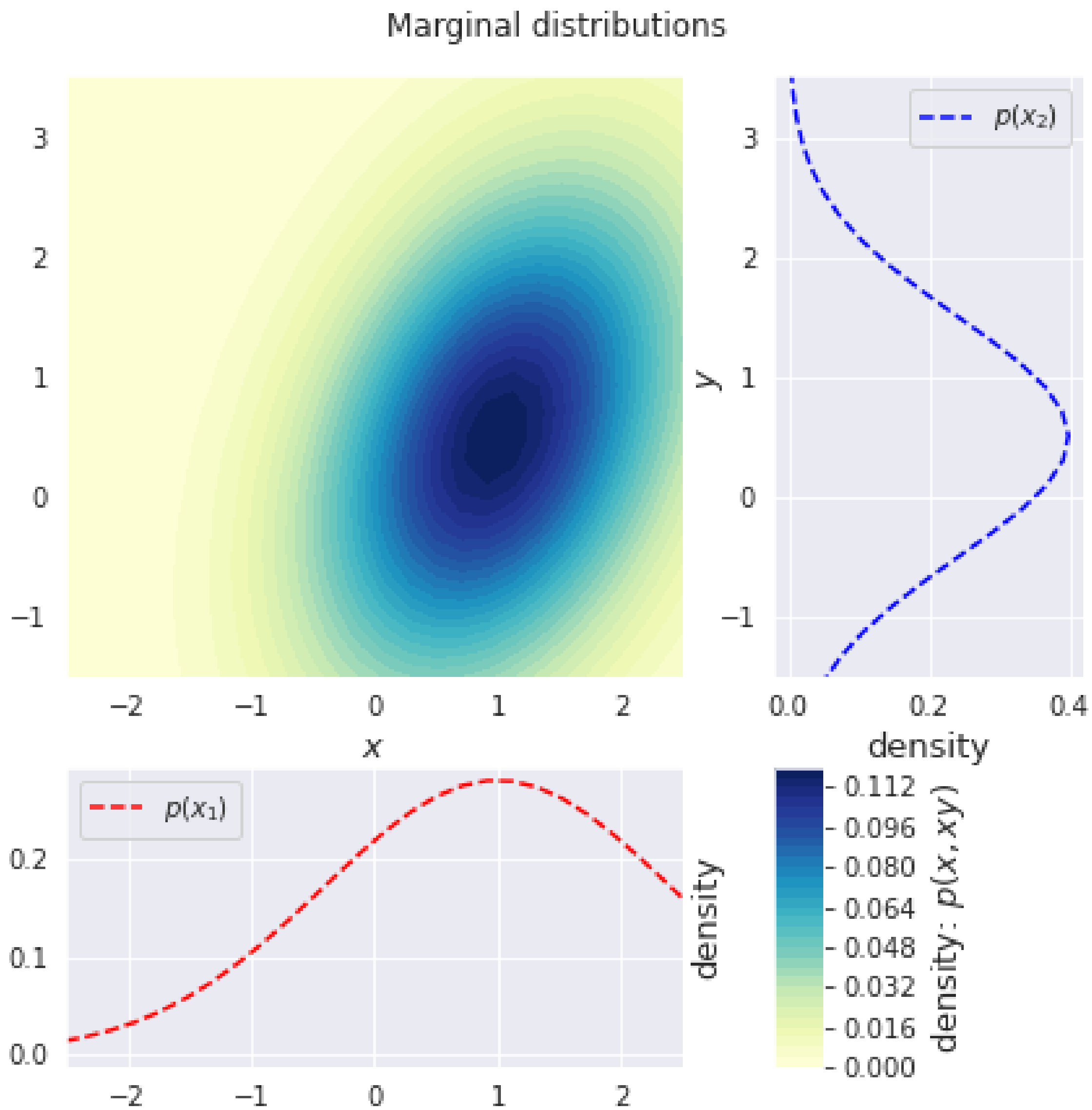

Figure 2: Marginal Distribution