

Assignment 1 Solutions

1 SVD and Eigendecomposition

1.1 SVD and Eigendecomposition Basics

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

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

- b) First method: We could just do the eigendecomposition of $A^T A$, 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 $A^T A$ 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 $A^T A$.
- c) As the singular values σ_i are the square roots of eigenvalues of $A^T A$, which are 6 and 1 we can get from b). Since $\text{rank}(A) = 2$, 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

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (1)$$

1.2 Geometric Interpretation of SVD

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

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

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

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

- b) First, V^T 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.

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

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

- b) The second order Taylor expansion is given by

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

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

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

- 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)

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

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

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

$$\vec{x} = (A^\top A)^{-1} A^\top \vec{b} \quad (8a)$$

$$= A^{-1}(A^\top)^{-1} A^\top \vec{b} \quad (8b)$$

$$= A^{-1} \vec{b} \quad (8c)$$

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,

$$A^\top \vec{b} = A^\top \vec{b}_1 + \vec{b}_2 \quad (9a)$$

$$= A^\top \vec{b}_1 \text{ since } \vec{b}_2 \in N(A^\top) \quad (9b)$$

$$= A^\top A \vec{x} \text{ for some } \vec{x}, \text{ since } \vec{b}_1 \in R(A) \quad (9c)$$

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.

3 Probability

3.1 Conditional Bayes' Rule

a)

By the Definition of Conditional Probability

$$p(x, y, z) = p(x, y|z)p(z) \Leftrightarrow \frac{p(x, y, z)}{p(z)} = p(x, y|z) \quad (10a)$$

By the General Product Rule

$$p(x, y, z) = p(x|y, z)p(y, z) = p(x|y, z)p(y|z)p(z) \quad (10b)$$

Combining (10a) and (10b)

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(x|y, z)p(y|z)p(z)}{p(z)} = p(x|y, z)p(y|z) \quad (10c)$$

b)

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

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

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

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

Noting

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

We have

$$p(x|z) = \frac{p(x, y|z)}{p(y|x, z)} \quad (11d)$$

Substituting (11d) into (11a)

$$\frac{p(x, y|z)}{p(x|z)} = \frac{p(x, y|z)p(y|x, z)}{p(x, y|z)} = p(y|x, z) \quad (11e)$$

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

$$p(\vec{x}; \mu, \Sigma) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^\top \Sigma^{-1}(\vec{x} - \vec{\mu})\right) \quad (12a)$$

$$= \frac{1}{2\pi(\sigma_1^2\sigma_2^2)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^\top \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\right) \quad (12b)$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2 - \frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2\right) \quad (12c)$$

$$= \frac{1}{2\pi\sigma_1^2\sigma_2^2} \exp\left(-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2\right) \exp\left(-\frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2\right) \quad (12d)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2\right) \quad (12e)$$

$$= p(x_1; \mu_1, \sigma_1^2)p(x_2; \mu_2, \sigma_2^2) \quad (12f)$$

where the fact that x_1 and x_2 are distributed normally is used from Eq. (12e) to (12f).

- 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.

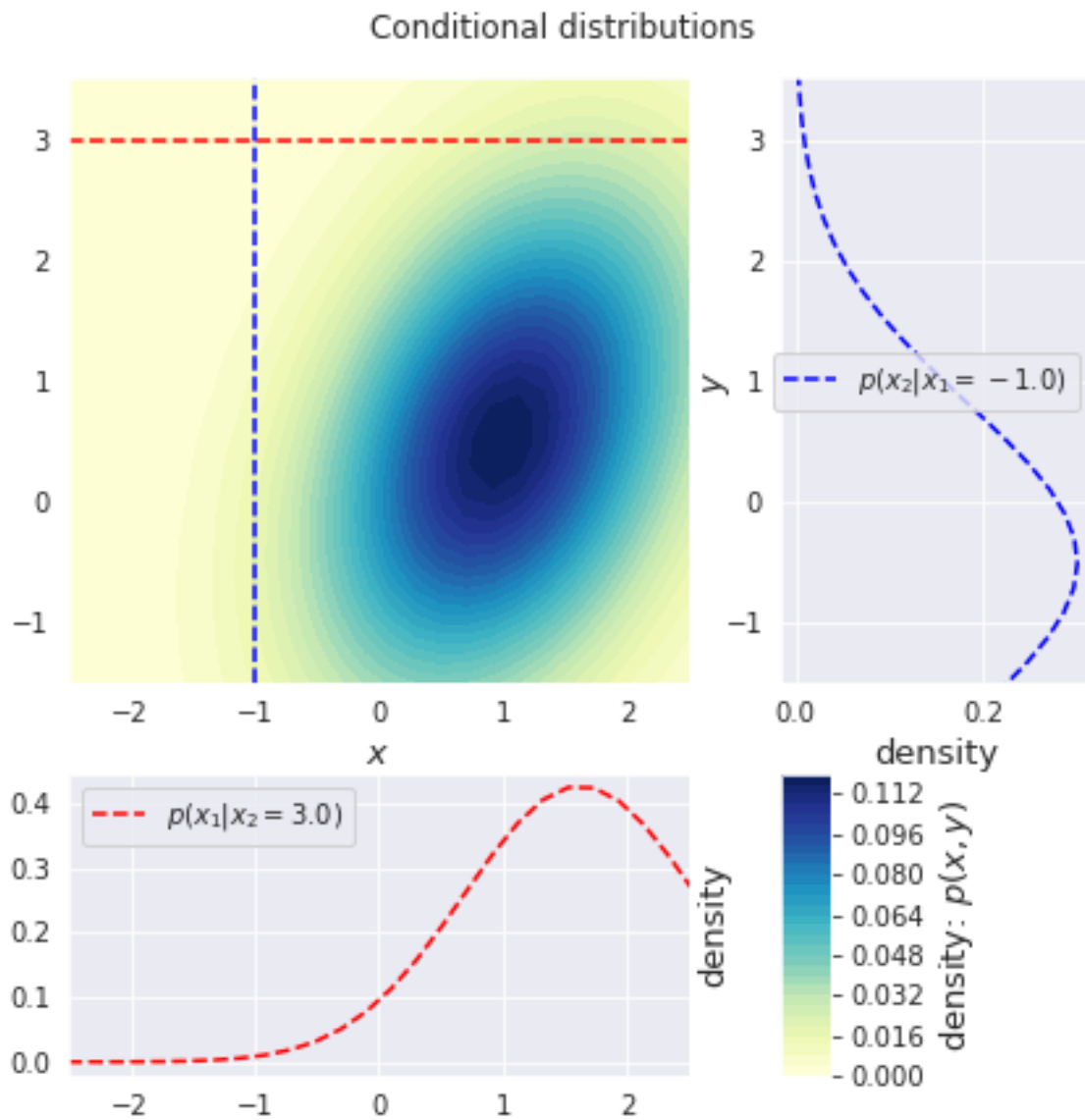


Figure 1: Conditional Distribution

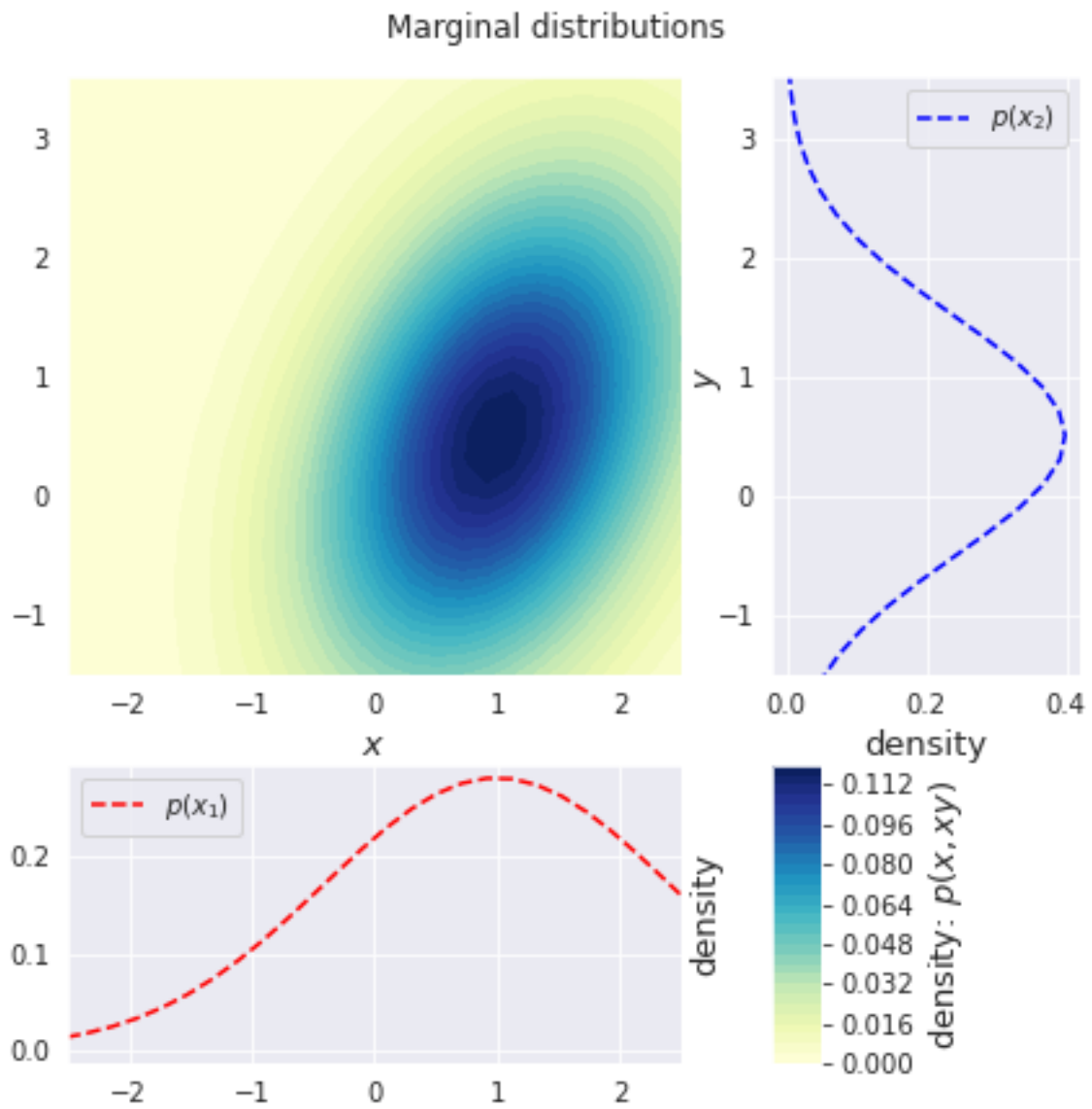


Figure 2: Marginal Distribution