6. Measuring the Spread of a Point Cloud

Review: Projection onto a Line

3/3 points (ungraded)

Let $\mathbf{u} \in \mathbb{R}^{\mathrm{d}}$ denote a unit vector (*i.e.*, $\sum_{i=1}^{d} \left(\mathbf{u}^i\right)^2 = 1$). In this unit, we will frequently refer to a unit vector as a **direction** , because we are primarily interested in the direction in which $\mathbf{u}$ is pointing.

In general, the **projection** of a vector $\mathbf{x} \in \mathbb{R}^d$ onto a **unit vector** $\mathbf{u}$ is defined to be

$$\mathrm{proj}_{\mathbf{u}}\mathbf{x} := (\mathbf{u} \cdot \mathbf{x})\,\mathbf{u}.$$

Note that if the vector onto which we project is not given as a unit vector but a vector, say $\mathbf{v}$, with length $\|\mathbf{v}\|$, then form the unit vector $\mathbf{u} = \dfrac{\mathbf{v}}{\|v\|}$ and apply the same formula as above: $\mathrm{proj}_{\mathbf{v}}\mathbf{x} = \left(\dfrac{\mathbf{v}}{\|\mathbf{v}\|} \cdot \mathbf{x}\right) \dfrac{\mathbf{v}}{\|\mathbf{v}\|} = \left(\dfrac{\mathbf{v} \cdot \mathbf{x}}{\|\mathbf{v}\|^2}\right) \mathbf{v}$.

In this problem, we set $d = 2$ and let $\mathbf{u} = \frac{1}{\sqrt{5}}(1,2)^T$. Suppose we have a data set consisting of three points given by

$$\begin{aligned} \mathbf{X}_1 &= (1,2)^T \\ \mathbf{X}_2 &= (3,4)^T \\ \mathbf{X}_3 &= (-1,0)^T. \end{aligned}$$

Find the vectors $\mathrm{proj}_{\mathbf{u}}\mathbf{X}_1$, $\mathrm{proj}_{\mathbf{u}}\mathbf{X}_2$, and $\mathrm{proj}_{\mathbf{u}}\mathbf{X}_3$. (Also plot them on a piece of paper.)

(Enter your answers as vectors, e.g., type **[3,2]** for the vector $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$).

$\mathrm{proj}_{\mathbf{u}}\mathbf{X}_1 =$ [1,2] ✔ **Answer:** [1,2]

$\mathrm{proj}_{\mathbf{u}}\mathbf{X}_2 =$ [2.2,4.4] ✔ **Answer:** [11/5,22/5]

$\mathrm{proj}_{\mathbf{u}}\mathbf{X}_3 =$ [-0.2,-0.4] ✔ **Answer:** [-0.2,-0.4]

Solution:

Note that $\mathbf{u}$ is a unit vector:

$$\|\mathbf{u}\|_2^2 = \frac{1}{5}(1^2 + 2^2) = 1.$$

By this fact and the given formula for the projection, we see that

$$\begin{aligned}
\mathrm{proj}_{\mathbf{u}}\mathbf{X}_1 &= (\frac{1}{\sqrt{5}}(1,2)^T \frac{1}{\sqrt{5}} \cdot (1,2)^T) \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\
\mathrm{proj}_{\mathbf{u}}\mathbf{X}_2 &= (\frac{1}{\sqrt{5}}(1,2)^T \cdot (3,4)^T) \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{11}{5} \\ \frac{22}{5} \end{pmatrix} \approx \begin{pmatrix} 2.2 \\ 4.4 \end{pmatrix} \\
\mathrm{proj}_{\mathbf{u}}\mathbf{X}_3 &= (\frac{1}{\sqrt{5}}(1,2)^T \cdot (-1,0)^T) \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -0.2 \\ -0.4 \end{pmatrix}
\end{aligned}$$

Remark 1: Observe that the point $(1,2)^T$ is already on the line that points in the direction of $\mathbf{u}$. Hence, the projection of this point onto $\mathbf{u}$ leaves this point fixed.

Remark 2: A geometric interpretation of $\mathrm{proj}_{\mathbf{u}}\mathbf{x}$ is given by the following. Consider the line L_1 that points in the direction of $\mathbf{u}$. Formally, this is defined as

$$L_1 := \{\lambda \mathbf{u} : \lambda \in \mathbb{R}\}.$$

Now consider the (unique) line L_2 that has the following properties:

- It passes through the endpoint of the vector $\mathbf{x}$,
- it passes through a point on the line L_1, and
- it is perpendicular to L_1.

The **intersection** of L_1 and L_2 is defined to be the endpoint of the vector $\mathrm{proj}_{\mathbf{u}}\mathbf{x}$.

You should compare this definition with the formula given for the projection and see, at least visually, that they give the same result.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Empirical Variance of a Data Set in a Given Direction

3/3 points (ungraded)

Consider the statistical set-up from the previous problem. In particular, recall that $\mathbf{u} = \frac{1}{\sqrt{5}}(1,2)^T$ and

$$\mathbf{X}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{X}_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \mathbf{X}_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

Observe that for $i = 1, 2, 3$, the number $\mathbf{u} \cdot \mathbf{X}_i$ (where $\mathbf{u}$ is a unit vector) gives the **signed distance** from the origin to the endpoint of the projection $\mathrm{proj}_{\mathbf{u}}\mathbf{X}_i$. By **signed distance** , we mean that $|\mathbf{u} \cdot \mathbf{X}_i|$ is the length of $\mathrm{proj}_{\mathbf{u}}\mathbf{X}_i$ and

$$\mathbf{u} \cdot \mathbf{X}_i > 0 \implies \mathbf{X}_i \text{ points approximately in the direction of } \mathbf{u}$$
$$\mathbf{u} \cdot \mathbf{X}_i < 0 \implies \mathbf{X}_i \text{ points approximately in the opposite direction of } \mathbf{u}$$

Compute the empirical variance of the data set

$$\mathbf{u} \cdot \mathbf{X}_1, \mathbf{u} \cdot \mathbf{X}_2, \mathbf{u} \cdot \mathbf{X}_3.$$

4.8 ✔ **Answer:** 4.80

Let $\mathbb{X}$ denote the matrix whose i-th row is $\mathbf{X}_i^T$.

Recall that $S = \frac{1}{3}\mathbb{X}^T \left(I_3 - \frac{1}{3}\mathbf{1}\mathbf{1}^T\right)\mathbb{X}$ denotes the empirical covariance matrix of our data set.

What is $\mathbf{u}^T S \mathbf{u}$?
(You are encouraged to use computational software.)

4.8 ✔ **Answer:** 4.80

Are your answers from part 1 and part 2 of this question the same?

- (●) Yes
- ( ) No

✔

Solution:

For the first part, since $\mathbf{u} = \frac{1}{\sqrt{5}}(1,2)^T$ is a unit vector, it suffices to just compute the dot products. Observe that

$$\mathbf{u} \cdot \mathbf{X}_1 \ = \sqrt{5} \approx 2.236$$

$$\begin{aligned}\mathbf{u}\cdot\mathbf{X}_2 &= \frac{11}{\sqrt{5}} \approx 4.919\\ \mathbf{u}\cdot\mathbf{X}_3 &= -\frac{1}{\sqrt{5}} \approx -0.447.\end{aligned}$$

The empirical mean is given by

$$\frac{1}{3\sqrt{5}}(5+11-1) \;=\; \sqrt{5} \approx 2.236.$$

Hence, the empirical variance is

$$\frac{1}{3}\left((\sqrt{5}-\sqrt{5})^2 + \left(\frac{11-5}{\sqrt{5}}\right)^2 + \left(\frac{-1-5}{\sqrt{5}}\right)^2\right) \;=\; \frac{72}{15} \;=\; 4.8$$

For the second part, we compute $\mathbf{u}^T S\mathbf{u}$.

$$\begin{aligned}\mathbf{u}^T\frac{1}{3}\mathbb{X}^T\left(I_3-\frac{1}{3}\mathbf{1}\mathbf{1}^T\right)\mathbb{X}\mathbf{u} &= \frac{1}{\sqrt{5}}(1,2)\,\frac{1}{3}\begin{pmatrix}1&3&-1\\2&4&0\end{pmatrix}\left(\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}-\frac{1}{3}\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}\right)\begin{pmatrix}1&2\\3&4\\-1&0\end{pmatrix}\frac{1}{\sqrt{5}}\begin{pmatrix}1\\2\end{pmatrix}\\ &= 4.8.\end{aligned}$$

We see that our answers to the first and second part are the same, so the correct answer to the third part is "Yes."

Remark: This problem and the previous one give an important geometric interpretation of the quantity $\mathbf{u}^T S\mathbf{u}$, where $\mathbf{u}$ is a fixed unit vector. Intuitively, we should think of this as describing the (empirical) variance of our data set in the direction of $\mathbf{u}$. The calculation above verifies this intuition on the data set above because we see that $\mathbf{u}^T S\mathbf{u}$ indeed gives the empirical variance of the *projected* data set $\mathbf{u}\cdot\mathbf{X}_1, \mathbf{u}\cdot\mathbf{X}_2, \mathbf{u}\cdot\mathbf{X}_3$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Variance of a Random Vector in a Given Direction

6/6 points (ungraded)

Let $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ where, for simplicity, $\mathbf{X} \in \mathbb{R}^2$ and hence $\Sigma \in \mathbb{R}^{2\times 2}$.

We can express $\mathsf{Var}(\mathbf{X}^1)$ as $\mathbf{u}^T\Sigma\mathbf{u}$ for some unit vector $\mathbf{u}$.

What is $\mathbf{u}$?

$\mathbf{u}^1 =$ 1 ✔ **Answer:** 1 , $\mathbf{u}^2 =$ 0 ✔ **Answer:** 0

Similarly, we can express $\mathsf{Var}(\mathbf{X}^2)$ as $\mathbf{v}^T\Sigma\mathbf{v}$ for some unit vector $\mathbf{v}$.

What is $\mathbf{v}$?

$\mathbf{v}^1 =$ 0 ✔ **Answer:** 0 , $\mathbf{v}^2 =$ 1 ✔ **Answer:** 1

Finally, we can express $\mathsf{Var}(\mathbf{X}^1 + \mathbf{X}^2)$ as $\mathbf{w}^T\Sigma\mathbf{w}$ for some vector $\mathbf{w}$.

What is $\mathbf{w}$?

$\mathbf{w}^1 =$ 1 ✔ **Answer:** 1 , $\mathbf{w}^2 =$ 1 ✔ **Answer:** 1

Solution:

Since $\mathbf{X}$ is centered (*i.e.*, it has mean $(0,0)^T$), we have that

$$\Sigma = \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right] = \begin{pmatrix} \mathbb{E}\left[\left(\mathbf{X}^1\right)^2\right] & \mathbb{E}\left[\mathbf{X}^1\mathbf{X}^2\right] \\ \mathbb{E}\left[\mathbf{X}^1\mathbf{X}^2\right] & \mathbb{E}\left[\left(\mathbf{X}^2\right)^2\right] \end{pmatrix}$$

In particular,

$$\Sigma_{11} = \mathsf{Var}\left(\mathbf{X}^1\right) = \mathbb{E}\left[\left(\mathbf{X}^1\right)^2\right]$$

and

$$\Sigma_{22} = \mathsf{Var}\left(\mathbf{X}^2\right) = \mathbb{E}\left[\left(\mathbf{X}^2\right)^2\right].$$

By inspection, we see that

$$\Sigma_{11} = (1,0)\,\Sigma(1,0)^T$$

and

$$\Sigma_{22} = (0,1)\,\Sigma(0,1)^T.$$

Therefore, $\mathbf{u} = (1,0)^T$ and $\mathbf{v} = (0,1)^T$.

For the last part, we take a different approach. Observe that

$$\begin{aligned} \mathbf{w}^T\Sigma\mathbf{w} &= \mathbf{w}^T\mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]\mathbf{w} \\ &= \mathbb{E}\left[\mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w}\right] \end{aligned}$$

$$= \mathbb{E}\left[(\mathbf{w}^T\mathbf{X})^2\right].$$

Note that $\mathbf{w}^T\mathbf{X} = \mathbf{w} \cdot \mathbf{X} \in \mathbb{R}$. Moreover,

$$\mathbb{E}\left[\mathbf{w}^T\mathbf{X}\right] = \mathbb{E}\left[\mathbf{w}^1\mathbf{X}^1 + \mathbf{w}^2\mathbf{X}^2\right] = 0$$

by the linearity of expectation. Hence,

$$\mathbf{w}^T\Sigma\mathbf{w} = \mathbb{E}\left[(\mathbf{w}\mathbf{X})^2\right] = \mathsf{Var}\left(\mathbf{w}^T\mathbf{X}\right)$$

for all vectors $\mathbf{w}$. Noting that $\mathbf{X}^1 + \mathbf{X}^2 = (1,1)^T\mathbf{X}$, we see that $\mathbf{w} = (1,1)^T$.

Remark: If we instead let $\mathbf{w} = \frac{1}{\sqrt{2}}(1,1)^T$ so that $\mathbf{w}$ is now a unit vector, then the quantity $\mathbf{w}^T\Sigma\mathbf{w}$ describes the variance of $\mathbf{X}$ in the direction of $\mathbf{w}$. More precisely, it is the variance of the random variable $w^T\mathbf{X} = \frac{1}{\sqrt{2}}(\mathbf{X}^1 + \mathbf{X}^2)$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

A Preview of Principal Component Analysis

1/1 point (ungraded)

This problem illustrates some of the main ideas behind principal component analysis, which will be explored in detail later in this lecture as well as the next lecture.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ denote a data set, and let $\mathbb{X}$ denote the matrix whose i-th row is $\mathbf{X}_i^T$. Let

$$S = \frac{1}{n}\mathbb{X}^T\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbb{X}$$

denote the empirical covariance matrix for this data set.

Consider the optimization problem

$$\text{argmax}_{\mathbf{u}:\|\mathbf{u}\|_2^2=1}\mathbf{u}^T S\mathbf{u}.$$

Let $\mathbf{u}^*$ denote the unit vector that maximizes $\mathbf{u}^T S\mathbf{u}$.

Which of the following is a correct interpretation of $\mathbf{u}^*$? (Choose all that apply.)

- [ ] $\mathbf{u}^*$ gives a direction such that the data set $\mathbf{u}^{*T}\mathbf{X}_1, \mathbf{u}^{*T}\mathbf{X}_2, \ldots, \mathbf{u}^{*T}\mathbf{X}_n$ is clustered closely together.
- [x] $\mathbf{u}^*$ is the direction that maximizes the empirical variance of the (projected) data points $\mathbf{u}^{*T}\mathbf{X}_1, \mathbf{u}^{*T}\mathbf{X}_2, \ldots, \mathbf{u}^{*T}\mathbf{X}_n$.
- [x] If $\mathbf{u}^{*T} S\mathbf{u}^*$ is very large, then if we project our data set onto the line spanned by $\mathbf{u}^*$, we expect the projected data set to be fairly 'spread out' (*i.e.*, the projected data set should have relatively large empirical variance).

✔

Solution:

We begin by examining the second and third choices, which are both correct. In the final bullet, we discuss the first choice, which is incorrect.

- The second choice is correct. As shown in a previous problem, for any unit vector $\mathbf{u} \in \mathbb{R}^d$, the empirical variance of the data set $\mathbf{u}^T\mathbf{X}_1, \ldots, \mathbf{u}^T\mathbf{X}_n \in \mathbb{R}$ is given by the quantity $\mathbf{u}^T S\mathbf{u}$. Hence, if we maximize $\mathbf{u}^T S\mathbf{u}$ over all unit vectors, then the maximizer $\mathbf{u}^*$ has the property that the empirical variance of

$$\mathbf{u}^{*T}\mathbf{X}_1, \mathbf{u}^{*T}\mathbf{X}_2, \ldots, \mathbf{u}^{*T}\mathbf{X}_n$$

is as large as possible.

- The third choice is also correct. For any unit vector $\mathbf{u} \in \mathbb{R}^d$, the data set

$$(\mathbf{u}^T\mathbf{X}_1)\,\mathbf{u}, (\mathbf{u}^T\mathbf{X}_2)\,\mathbf{u} \ldots, (\mathbf{u}^T\mathbf{X}_n)\,\mathbf{u}$$

is formed by projecting the data points $\mathbf{X}_1, \ldots, \mathbf{X}_n$ onto the line spanned by the vector $\mathbf{u}$. Note that for all i, the number $\mathbf{u}^T\mathbf{X}_i$ denotes the distance from the origin of the projection of $\mathbf{X}_i$ onto the line in the direction of $\mathbf{u}$.

By the previous bullet, $\mathbf{u}^*$ maximizes the empirical variance $\mathbf{u}^T S\mathbf{u}$ of $\mathbf{u}^T\mathbf{X}_1, \ldots, \mathbf{u}^T\mathbf{X}_n$ over all unit vectors $\mathbf{u}$. Hence, we expect the points

$$(\mathbf{u}^{*T}\mathbf{X}_1)\,\mathbf{u}^*, (\mathbf{u}^{*T}\mathbf{X}_2)\,\mathbf{u}^* \ldots, (\mathbf{u}^{*T}\mathbf{X}_n)\,\mathbf{u}^*$$

to be very spread out if the $\mathbf{u}^T S\mathbf{u}$ is very large.

- The first choice "$\mathbf{u}^*$ gives a direction in which the data set is clustered closely together." is incorrect. As explained in the above two bullets, the unit vector $\mathbf{u}^*$ should, intuitively speaking, denote a direction where our data set is fairly spread out and, hence, **not** clustered closely together.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: (Optional) Unit 8 Principal component analysis:(Optional) Preparation Exercises for Principal Component Analysis / 6. Measuring the Spread of a Point Cloud

Add a Post

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.

✖