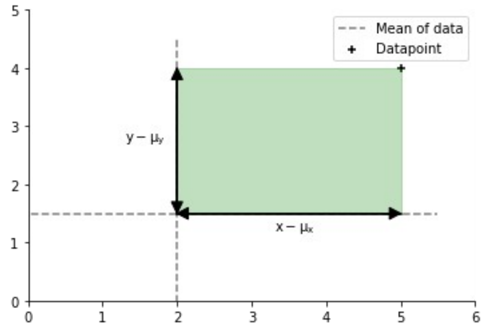
• En esta pregunta, veremos un conjunto de datos bidimensional. $\mathcal{D} = \{ \mathbf{x}_i \}_{yo=1}^{norte}$ con $norte$ muestras cada muestra $\mathbf{X}_i$ en el conjunto de datos es un vector bidimensional con coordenadas X, y, es decir, la primera componente del vector se denota por X y el otro por y.

1 / 1 punto

La covarianza entre dos variables aleatorias *escalares es*

$$\mathrm{c\,o\,v}\,[\,x\,,y] = y\,[\,(\,\mathrm{x} - metro_{\mathrm{X}})(\mathrm{y} - metro_{\mathrm{y}})] \approx \frac{1}{norte} \sum_{yo=1}^{norte} (\,\mathrm{X} - metro_{\mathrm{X}})(\mathrm{y} - metro_{\mathrm{y}}).$$

En la fórmula de la covarianza, podemos pensar en cada multiplicación individual como el cálculo de un área, un rectángulo con lados $\mathrm{X} - metro_{\mathrm{X}}$ y $\mathrm{y} - metro_{\mathrm{y}}$.



Para este punto de datos, un aumento en X de la media está relacionado con un aumento en y. Donde $\mathrm{X} - metro_{\mathrm{X}}$ y $\mathrm{y} - metro_{\mathrm{y}}$ tienen el mismo signo, la contribución a la covarianza es positiva y de color verde, mientras que si los signos son opuestos será negativa y de color rojo. En otras palabras, verde significa que X y y están correlacionados positivamente, mientras que el rojo significa que están correlacionados negativamente.

La suma total de áreas, dividida por el número de puntos $norte$, será el valor de la covarianza.

Ejecute el código una vez para ver esto, luego elimine el comentario de la línea que mostrará los rectángulos y ejecútelo de nuevo.

```
1  # EJECUTE EL CÓDIGO UNA VEZ, LUEGO DESCOMENTE LA LÍNEA 29 PARA VISUALIZAR LA COV
2
3  higo, hacha = plt.subplots()
4
5  #Elija una matriz eliminando el # delante de la  palabra "datos" a continuación
6  #Para cambiar, vuelve a poner el # y borra otro
7
8  #Aleatorio:
9  datos = np.array([[ 1 , 2 ],[ 5 , 4 ],[ -2 , -3 ],[ 4 , -2 ],[ 2 , 3 ],[ 8 , -9
10
11 #Línea recta:
12 datos = np.arreglo([[ 1 , 1 ],[ -3 , -3 ],[ 2 , 2 ],[ 7 , 7 ]])
13
14 #Q1: cuadrado
15 datos = np.arreglo([[ 0 , 0 ],[ 4 , 4 ],[ 0 , 4 ],[ 4 , 0 ]])
16
17 #¡Siéntete libre de ingresar tu propia matriz o modificar las anteriores!
18
19 # Primero calcule la media con la función NumPy np. significar().
20 # El primer argumento es el conjunto de datos y "eje" especifica la dirección
21 # La varianza en 1D se puede calcular de manera similar con  np.var()
22 datos_medios = np.medio(datos, eje= 0 )
23 create_plot(data)  #que también agrega varianzas 1d
24
25 área= 0
26 media = media_datos
27
28 para  i  en rango ( len (datos)):
29     show_rectangle(media, datos[i])
30     # y un cálculo que suma (o resta)
31     # el valor del área a nuestro valor de la covarianza :
32     área += calcular_área(media, datos[i])
33
34 plt.show()
```

Ejecutar

Restablecer

Las líneas discontinuas se encuentran en la media del conjunto de datos. Las líneas azules representan la magnitud de la varianza de los componentes x (horizontal) e y (vertical) del conjunto de datos.

Si el rojo y el verde se equilibran, la covarianza será 0. De lo contrario, el signo de la covarianza dará una dirección en la que los puntos parecen estar correlacionados.

Que es $cov(x, y)$para el conjunto de datos en la matriz con la etiqueta "Q1: cuadrado"? ¿Es lo que esperarías de la trama?

0

Correcto

Correct! Since the points are evenly distributed around the mean they balance out and there is no way to determine a direction of correlation.

2. The covariance matrix is given by

1 / 1 punto

$$\begin{bmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) \end{bmatrix} = \begin{bmatrix} \mathrm{var}(x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{var}(y) \end{bmatrix}$$

Compute the covariance matrix for the following dataset

$$D = \{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \end{bmatrix} \}$$

Here, every column vector represents a data point.

Do the exercise using pen and paper. You can check if your answer makes sense with this codeblock.

```
data = np.array([[1,2],[5,4]])

mean_data = np.mean(data, axis=0)
create_plot(data)

area=0
mean = mean_data

for i in range(len(data)):
    show_rectangle(mean, data[i])
    area += calculate_area(mean, data[i])

plt.show()
```

Ejecutar

Restablecer

- ○ $\begin{bmatrix} 2 & 2 \\ 4 & 1 \end{bmatrix}$
- ◉ $\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$
- ○ $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

Correcto

Good job!

3. Consider a data set D with covariance matrix $\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$.

1 / 1 punto

What is the covariance matrix if we multiply every vector in D by 2?

Run the codeblock below to observe what happens to the example in Question 2 (this will NOT give you the answer to this question but might aid intuition).

```
data = np.array([[1,2],[5,4]])
data *= 2
#Uncomment the line above to multiply by 2 and run again

mean_data = np.mean(data, axis=0)
create_plot(data)

area=0
mean = mean_data

for i in range(len(data)):
    show_rectangle(mean, data[i])
    area += calculate_area(mean, data[i])

plt.show()
```

Ejecutar

Restablecer

- $\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$
- $\begin{bmatrix} 12 & 8 \\ 8 & 16 \end{bmatrix}$
- $\begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$
- $\begin{bmatrix} 16 & 8 \\ 8 & 12 \end{bmatrix}$

Correcto

Yes, every element in the covariance matrix is multiplied by 4.

4. Considere el conjunto de datos $D = \{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 7 \\ 4 \end{bmatrix} \}$ con matriz de covarianza $\begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}$.

1 / 1 punto

Calcule la nueva matriz de covarianza cuando sumamos $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ a cada elemento en D.

Ejecute el bloque de código a continuación para observar qué sucede con el ejemplo de la Pregunta 2 cuando se agrega un vector a cada punto (esto NO le dará la respuesta, pero podría ayudar a la intuición).

```
datos = np.matriz([[ 1 , 2 ],[ 5 , 4 ]])
datos += [ 2 , 2 ]
#Descomenta la línea anterior después de la primera ejecución para agregar [2,2

datos_medios = np.medio(datos, eje= 0 )
create_plot(datos)

área= 0
media = media_datos

para  i  en rango ( len (datos)):
    show_rectangle(media, datos[i])
    área += calcular_área(media, datos[i])

plt.show()
```

Ejecutar

Restablecer

- $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- [x] $\begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}$

- [ ] $\begin{bmatrix} 11 & 5 \\ 5 & 3 \end{bmatrix}$

Correcto
Bien hecho. La covarianza no cambiará.

5. Estamos viendo un conjunto de datos. D donde cada elemento en D consiste en un X y coordinar. La matriz de covarianza de datos está dada por

1 / 1 punto

$$\begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

¿Cuál de las siguientes afirmaciones es correcta?

- [x] X y están correlacionados positivamente, es decir, cuando X aumenta entonces y aumenta en promedio, y viceversa.

- [ ] X y están correlacionados negativamente, es decir, cuando X aumenta entonces y disminuye en promedio, y viceversa.

- [ ] X y no están correlacionados, es decir, cuando X aumenta entonces y no cambia en promedio (y viceversa).

Correcto
¡Bien hecho!