CS 2810 April 8

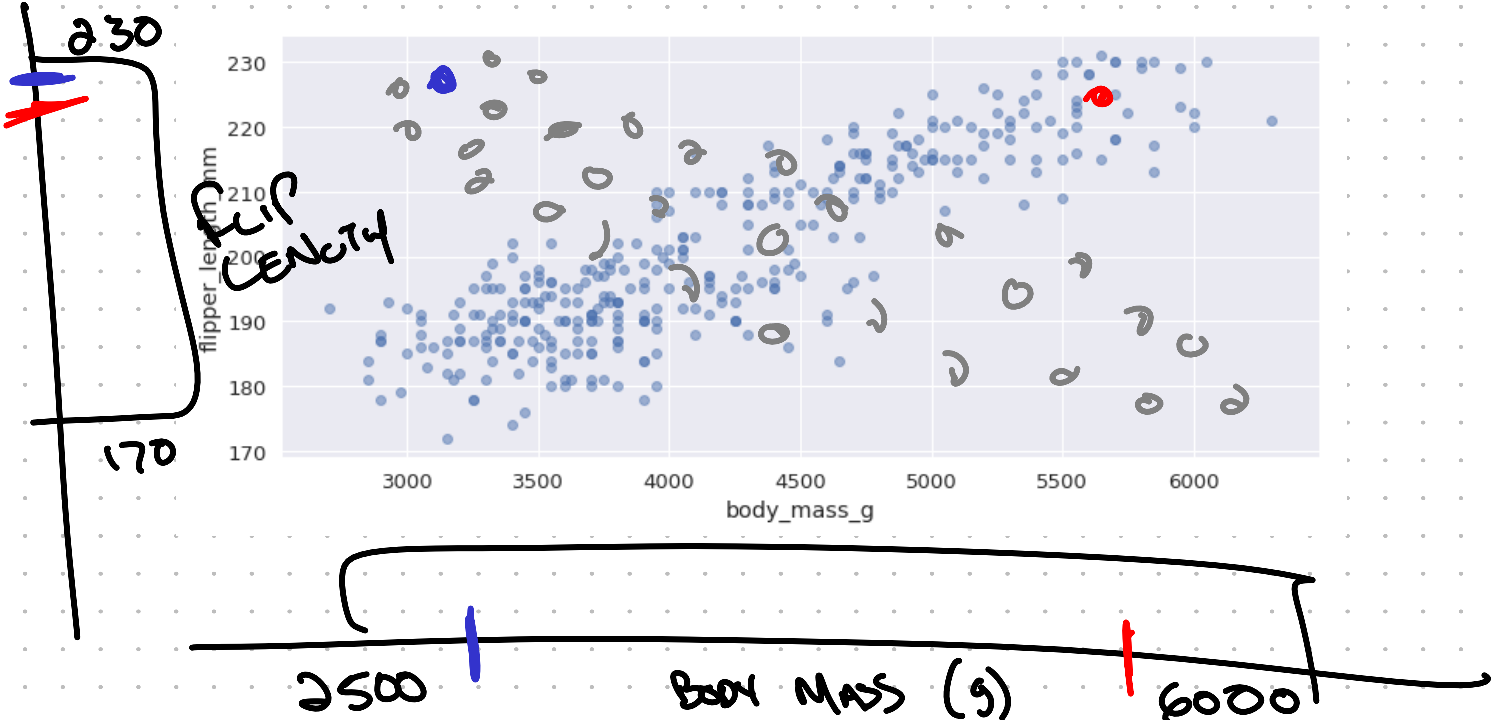
Admin:
- schedule update: we'll do a finals review in class & move up mini project day 2

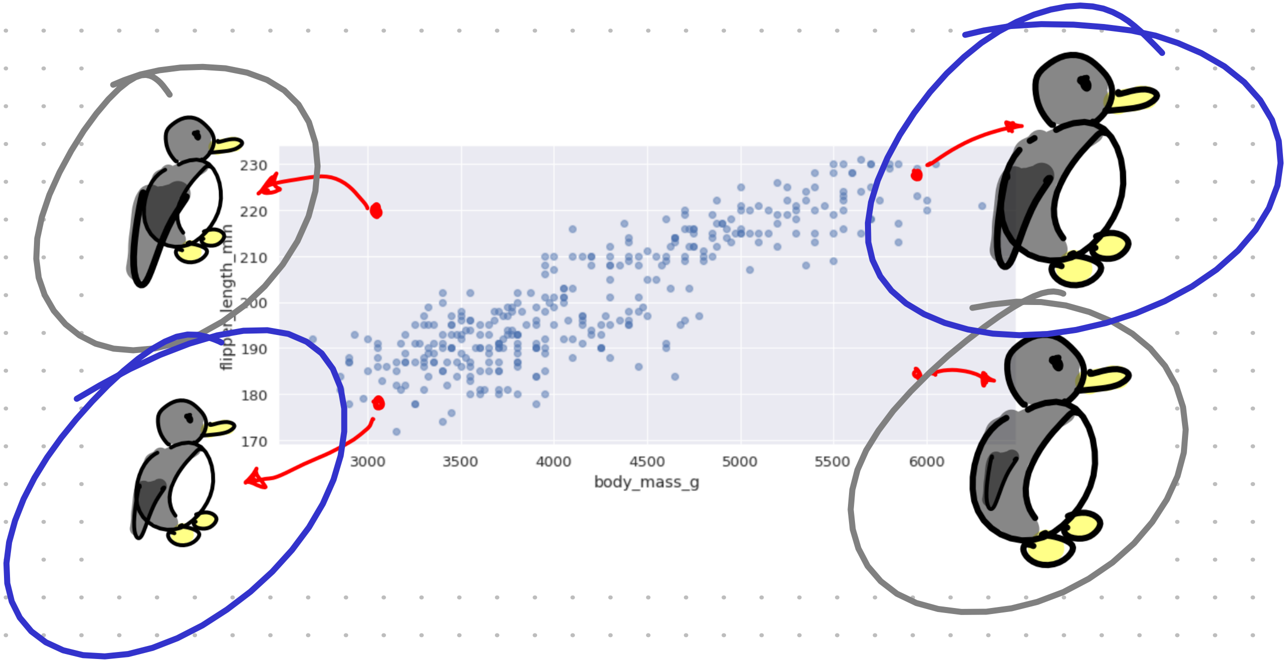
Content:
Covariance

Covariance Matrix

(covariance and correlation are tightly intertwined, we'll touch on
correlation in the next lesson ... please save correlation questions until then, thanks!)

1. Why would one want to generate a "fake" sample?
2. Generating a "fake" penguin from marginal distributions





230
220
210
200
190
180
170
3000
3500
4000
4500
5000
5500
6000
body_mass_g

WE NEED A WAY OF QUANTIFYING

HOW (BODY_MASS) VARIES WITH (FLIPPER_LENGTH)

X Y

(. . . COVARIANCE)

Covariance: Inuition (How two values vary together)

The behavior between any two values x and y can be summarized in one of the three ways:

1. as x gets larger y typically gets larger too
   - ex:
     - `x=temp on some day`
     - `y=number of people on the beach on the same day`
   - covariance & correlation is positive
2. as x gets larger y typically doesn't get larger or smaller
   - ex:
     - `x=individual's favorite number`
     - `y=number of hot dogs that individual has eaten in their lifetime`
   - covariance & correlation is zero
3. as x gets larger, y typically gets smaller
   - ex:
     - `x=average speed of driver on 10 mile commute`
     - `y=average commute time of driver on 10 mile commute`
   - covariance & correlation are negative

Joint Distribution + Marginalization (Review)

	Circle $C=0$	$C=1$
Shaded $S=0$	$1/3$	0
$S=1$	$1/2$	$1/6$

What is $P(S=0)$?

$$P(S=0 \;\; C=0) + P(S=0 \;\; C=1)$$

$$= 1/3 + 0 = 1/3$$

Joint Distribution

$P(S,C)$ describes prob of S,C outcome pairs

(Go back)

Marginal Distribution

Prob of just one variable

$$P(S) = \sum_{c} P(S \;\; C)$$

All outcomes c

A JOINT DISTRIBUTION IMPOSES MARGINAL DISTRIBUTION

THERE CAN EXIST MANY JOINT DISTRIBUTIONS WHICH HAVE SAME MARGINALS

→ HOW DO X, Y VARY TOGETHER? (... COV)

Observations of x, y must be paired for a joint distribution / covariance to be defined.

1. on the same **day**, we observe temp (x) & beach population (y)
2. on the same **individual**, we observe favortie number (x) & hot dogs eaten (y)
3. on the same **driver**, we observe speed (x) & commute time (y)

If we don't observe the data in pairs, correlation / covariance is not defined:

- `x=an individual's favorite number`
- `y=the temperature on a given day`

x is observed per individual
y is observed per day

... there isn't a way to "pair" every x with a y!

Covariance

How do R.V. X and Y vary together?

$$\sigma_{XY}^2 = \operatorname{Cov}(X,Y) = E\left[\underbrace{(X-E[X])(Y-E[Y])}_{\text{can be negative}}\right]$$

Note:

$$E\left[\underbrace{(X-E[X])}_{\text{always non negative}}{}^2\right] = \operatorname{Var}(X) = \sigma_X^2$$

Example: Compute cov(x,y) and write one sentence which explains its sign.

X \ Y	0	1
0	.4	.1
1	.1	.4

JOINT DISTRIBUTION

$$\mathrm{cov}(x,y) = E\left[(x - E[x])(y - E[y])\right]$$

$$E[x] = \sum_x x P(x) = 0(.5) + 1(.5) = .5$$

$$P(x=0) = P(x=0, y=0) + P(x=0, y=1) = .4 + .1 = .5$$

Example: Compute cov(x,y) and write one sentence which explains its sign.

$$E[x] = E[y] = .5$$

$$\begin{array}{cc|cc} & & y & \\ & & 0 & 1 \\ \hline x & 0 & .4 & .1 \\ & 1 & .1 & .4 \end{array}$$

JOINT DISTRIBUTION

$$\text{cov}(xy) = E\Big[\big(x - E[x]\big)\big(y - E[y]\big)\Big]$$

$$\text{cov}(xy) = \sum_{xy} p(x,y)\,\big(x - E[x]\big)\big(y - E[y]\big)$$

$$= .4(0-.5)(0-.5) + .1(0-.5)(1-.5) + .1(1-.5)(0-.5) + .4(1-.5)(1-.5) = .15$$

cov is positive since when x gets bigger
y does too (similarly, when y bigger x too!)

SAMPLE COVARIANCE (ESTIMATING COV)

$$\sigma^2_{xy} = \mathrm{Cov}(X,Y) = E\left[(X-E[X])(Y-E[Y])\right]$$

GROUND TRUTH (REQUIRES DISTRIBUTION)

$$\hat{\sigma}^2_{xy} = \frac{1}{N-1}\sum_{x,y}(X-\bar{X})(Y-\bar{Y})$$

ESTIMATOR

BESSEL => UNBIASED

In Class Assignment 1:

$$\hat{\sigma}_{xy}^2 = \frac{1}{N-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Compute the sample covariance from the x,y samples given below. Write one sentence which interprets its sign.

X	1	2	3	4	5
Y	1	4	9	16	20

$\bar{x} = 3$

$\bar{y} = 10$

WE OBSERVE $x=1$ WHILE $y=1$

$$\hat{\sigma}_{xy}^2 = \frac{1}{5-1}\left[(1-3)(1-10) + (2-3)(4-10) + (3-3)(9-10) + (4-3)(16-10) + (5-3)(20-10)\right]$$

$$= \frac{1}{4}\left(18 + 6 + 0 + 6 + 20\right) = 12.5$$

In Class Assignment 1:

$$\hat{\sigma}_{xy}^2 = \frac{1}{N-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Compute the sample covariance from the x,y samples given below. Write one sentence which interprets its sign.

x	6	5	5	4
y	1	2	2	3

$\bar{x} = 5$

$\bar{y} = 2$

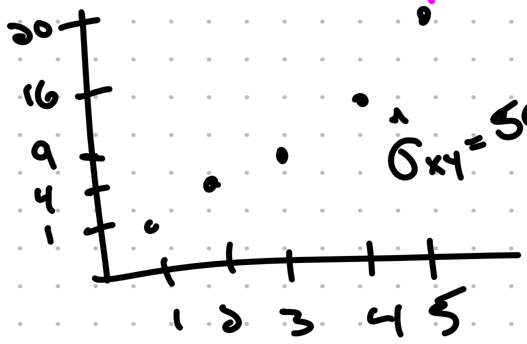
$$\hat{\sigma}_{xy}^2 = \frac{1}{4-1}\Big[(6-5)(1-2) + (5-5)(2-2) + (5-5)(2-2) + (4-5)(3-2)\Big]$$

$$= 1/3\,(\,-1 + 0 + 0 + -1\,) = -2/3$$

INTERPETTING SIGN OF COV

X	1	2	3	4	5
Y	1	4	9	16	20

+ COV ⇒ WHEN X IS LARGER Y GETS LARGER



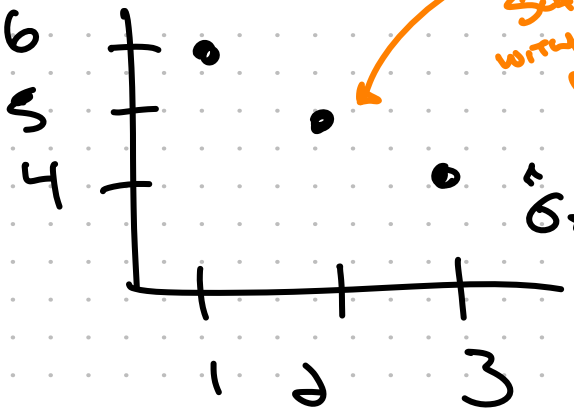
$\hat{\sigma}_{xy} = 50/4$

(NOT TO SCALE)

- COV ⇒ WHEN X IS LARGER Y IS SMALLER

X	6	5	5	4
Y	1	2	2	3

⚠ SCATTER LIES WITH REDUNDANT POINT



$\hat{\sigma}_{xy} = -2/3$

Anatomy of a covariance matrix:

Given a vector of random variables, we can describe the covariance of every pair of variables with a covariance matrix:

$$\vec{X} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \mathrm{cov}(x_0, x_0) & \mathrm{cov}(x_0, x_1) \\ \mathrm{cov}(x_1, x_0) & \mathrm{cov}(x_1, x_1) \end{bmatrix}$$

VAR(x_0)

VAR(x_1)

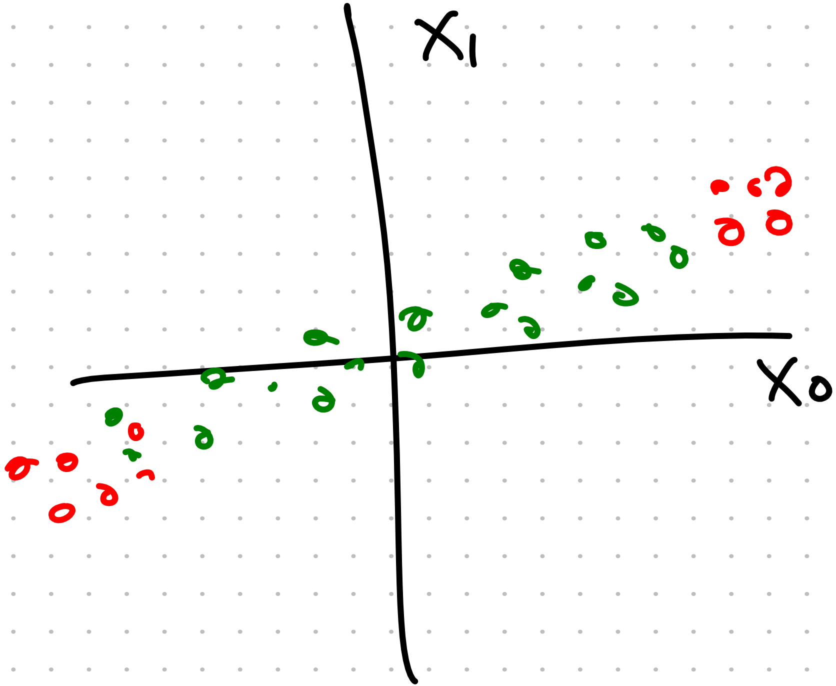
VECTOR OF RANDOM VARIABLES

COV MATRIX (NOT SUMMATION)

→ DIAGONAL TELLS US HOW FAR EACH SAMPLE IS FROM MEAN

→ COV MATRICES ARE SYMMETRIC

$$\Sigma = \begin{bmatrix} 10 & 2 \\ 2 & 1 \end{bmatrix}$$



$$\vec{X} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \mathrm{cov}(x_0 x_0) & \mathrm{cov}(x_0 x_1) & \mathrm{cov}(x_0 x_2) \\ \mathrm{cov}(x_1 x_0) & \mathrm{cov}(x_1 x_1) & \mathrm{cov}(x_1 x_2) \\ \mathrm{cov}(x_2 x_0) & \mathrm{cov}(x_2 x_1) & \mathrm{cov}(x_2 x_2) \end{bmatrix}$$

OBSERVATION 1 $\quad \mathrm{COV}(X, X) = \mathrm{VAR}(X)$

$$\mathrm{COV}(X, X) = E\Big[\big(X - E[X]\big)\big(X - E[X]\big)\Big]$$

$$= E\Big[\big(X - E[X]\big)^2\Big] = \mathrm{VAR}(X)$$

(makes sense why we call it "co"-variance, right?)

Observation 2

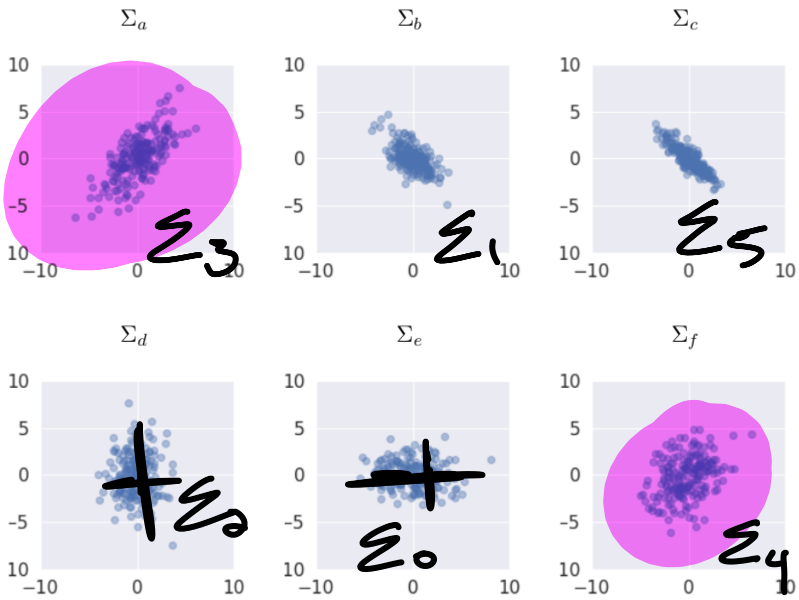
$$\mathrm{cov}(X\,Y) = \mathrm{cov}(Y\,X)$$

$$\mathrm{cov}(X\,Y) = E\Big[\big(X - E[X]\big)\big(Y - E[Y]\big)\Big]$$

$$= E\Big[\big(Y - E[Y]\big)\big(X - E[X]\big)\Big] = \mathrm{cov}(Y\,X)$$

In Class Assignment 2: Covariance Matching

Match each scatter plot on the right with its associated covariance below.



$$\Sigma_0 = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 2 & -1.3 \\ -1.3 & 2 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$$

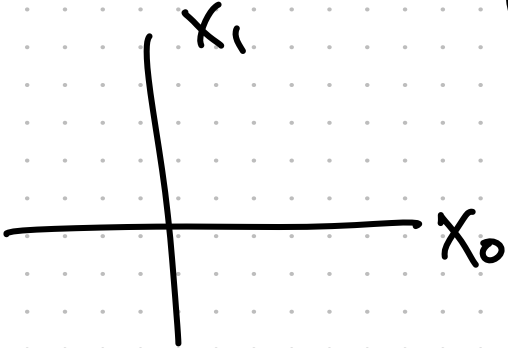
$$\Sigma_3 = \begin{bmatrix} 4 & 3 \\ 3 & 5 \end{bmatrix}, \quad \Sigma_4 = \begin{bmatrix} 4 & 1 \\ 1 & 5 \end{bmatrix}, \quad \Sigma_5 = \begin{bmatrix} 2 & -1.7 \\ -1.7 & 2 \end{bmatrix},$$

HORZ SPREAD

HOW DO X,Y VARY TOGETHER?

$$\Sigma = \begin{bmatrix} VAR(X_0) & COV(X_0 X_1) \\ COV(X_0 X_1) & VAR(X_1) \end{bmatrix}$$

VERT SPREAD



ICA 3 (if time) Interpretting a covariance matrix:

The covariance matrix to the right describes covariances between four penguin features.

1. Which penguin feature varies the most? (Can you compare across units?)
2. Give an intuition for what the -748 is saying in the matrix, does this make sense to you?

	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g
bill_length_mm	29.906333	-2.462091	50.058195	2595.623304
bill_depth_mm	-2.462091	3.877888	-15.947248	-748.456122
flipper_length_mm	50.058195	-15.947248	196.441677	9852.191649
body_mass_g	2595.623304	-748.456122	9852.191649	648372.487699

$$\bar{X} = \begin{bmatrix} 3 \\ 2 \\ 5 \\ 5 \end{bmatrix}$$

-748 IMPLIES

AS BODY MASS ↑

BILL DEPTH ↓

$X' = X \cdot C$

MASS IN TONS

MASS IN GRAMS

VERY SMALL NUMBER

$\mathrm{VAR}(X') = \mathrm{VAR}(X \cdot C)$

$= C^2 \, \mathrm{VAR}(X)$

EVEN SMALLER NUMBER

Use these scatters to validate your thinking on ICA 3.

Try to form your intuition for the cov matrix alone, before using these plots.