EECS 245, Spring 2026

LEC 13 Singular Value Decomposition, Principal Components Analysis

→ Read: Ch. 10

Agenda

Ch. 10

- Singular value decomposition SVD

$$X = U \Sigma V^T$$

  - where did it come from?
  - what can it be used for?

- Principal components analysis PCA: a cool application of the SVD

Announcements

- Lab 11 solutions up
- Deadlines:
  - HW 10 today
  - HW 11 Sunday
  - Lab 12 Monday
- Final Exam on Wednesday, 8-10AM
- After HW 10, fill out the End-of-Semester Survey and Official Evals for 1% EC

Recap: what is the singular value decomposition (SVD)?

X $n \times d$, not necessarily square

$$X_{n\times d} = U_{n\times n}\,\Sigma_{n\times d}\,V_{d\times d}^T$$

orthogonal!
$U^TU = UU^T = I_{n\times n}$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix}$$

orthogonal!
$V^TV = VV^T = I_{d\times d}$

$\underbrace{\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_r > 0}$

each σ_i is a singular value, # non-zero σ_i's $= \text{rank}(X)$

Equation that relates cols of U to cols of V:

$$X \vec{v}_i = \sigma_i \vec{u}_i$$

i^{th} column of U

$\vec{u}_i \in \mathbb{R}^n$

i^{th} column of V

$\vec{v}_i \in \mathbb{R}^d$

$$X = \begin{bmatrix} 3 & 2 & 5 \\ 2 & 3 & 5 \\ 2 & -2 & 0 \\ 5 & 5 & 10 \end{bmatrix} \qquad X \underset{\vec{v}_1}{\begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}} = \underset{\sigma_1}{15} \underset{\vec{u}_1}{\begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ 0 \\ 2/\sqrt{6} \end{bmatrix}}$$

$$X = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{3\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{2}{3} \\ \frac{1}{\sqrt{6}} & -\frac{1}{3\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & 0 & \frac{1}{3} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}}_{V^T}$$

Where does the SVD come from?

$$X = U \Sigma V^T$$

X is not square, so it doesn't have any eigen-things.

But $X^T X$ and XX^T are both square!

They are also symmetric, thus the spectral theorem tells us they are diagonalizable with orthogonal eigenvectors.

Recap: diagonalizable matrices and the spectral theorm

- If A $(n \times n)$ diagonalizable, then exists

$$A = V \Lambda V^{-1}$$

$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \end{bmatrix}$

cols are eigenvectors of A

- Spectral theorem says if A symmetric $(A = A^T)$, then

$$A = Q \Lambda Q^T$$

$Q^T Q = I$, so $Q^{-1} = Q^T$

cols are eigenvectors of A

$$X = U\Sigma V^T$$

not summation !

$$X^TX = (U\Sigma V^T)^T U\Sigma V^T$$

$$= V\Sigma^T \underbrace{U^T U}_{I} \Sigma V^T$$

$$= V\Sigma^T\Sigma V^T$$

looks like $Q\Lambda Q^T$

$$XX^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma \overbrace{V^T V}^{I} \Sigma^T U^T$$

$$= U\Sigma\Sigma^T U^T$$

also looks like $Q\Lambda Q^T$

$\Sigma \quad n \times d$

$\Sigma^T \Sigma \quad d \times d$

$\Sigma \Sigma^T \quad n \times n$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{(example)}$$

e.g. $X \quad 5 \times 3$, rank $= 2$

$$\Sigma^T \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

square, diagonal

$$\Sigma \Sigma^T = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & 0 & \\ & & 0 & & \\ & 0 & & 0 & \\ & & & & 0 \end{bmatrix}$$

square, diagonal

$$X = U \Sigma V^T$$

$$X^T X = V \Sigma^T \Sigma V^T = V \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \end{bmatrix} V^T$$

→ $X^T X$ is a symmetric matrix!

→ $V(\Sigma^T \Sigma) V^T$ is just the spectral decomposition

→ columns of V are eigenvectors of $X^T X$

→ $\sigma_i^2 = \lambda_i$ of $X^T X$

→ $\sigma_i = \sqrt{\lambda_i \text{ of } X^T X}$

$X^T X$'s eigenvalues are ≥ 0, since $X^T X$ is positive semi-definite!

number of non-zero σ_i's $= \operatorname{rank}(X) = r$

$$X = \underbrace{\begin{bmatrix} \overbrace{\begin{matrix} | & \cdots & | \\ \vec{u}_1 & \cdots & \vec{u}_r \\ | & \cdots & | \end{matrix}}^{\text{basis for colsp}(X)} & \overbrace{\begin{matrix} | & \cdots & | \\ \vec{u}_{r+1} & \cdots & \vec{u}_n \\ | & \cdots & | \end{matrix}}^{\text{basis for nullsp}(X^T)} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \left.\begin{matrix} - \; \vec{v}_1^T \; - \\ \vdots \\ - \; \vec{v}_r^T \; - \end{matrix}\right\} \text{basis for colsp}(X^T) \\ \left.\begin{matrix} - \; \vec{v}_{r+1}^T \; - \\ \vdots \\ - \; \vec{v}_d^T \; - \end{matrix}\right\} \text{basis for nullsp}(X) \end{bmatrix}}_{V^T}$$

$\vec{u}_1, \ldots, \vec{u}_n$ eigenvectors of XX^T

$\vec{v}_1, \ldots, \vec{v}_d$ (columns of V) eigenvectors of X^TX

One example SVD

Find SVD of X

$\text{rank}(X) = 2$

$$X = \begin{bmatrix} 3 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 25 & 3 & 4 \\ 3 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

$\rightarrow$ eigenvalues:

- since $\text{rank}(X) = \text{rank}(X^T X) = 2$, $X^T X$ not invertible, so 0 is an eigenvalue

$\rightarrow$ what are the other 2?

26
1
0

$$XX^T = \begin{bmatrix} 10 & 12 \\ 12 & 17 \end{bmatrix}$$

$\rightsquigarrow$ $\lambda_1 \lambda_2 = 170 - 144 = 26$

$\lambda_1 + \lambda_2 = 27$

$\Rightarrow$ $\lambda_1 = 26$

$\lambda_2 = 1$

$\sigma_1 = \sqrt{26}$

$\sigma_2 = 1$

$$\Sigma = \begin{bmatrix} \sqrt{26} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

eigenvalue 26: $\vec{v}_1 = ?$

$$X^T X = \begin{bmatrix} 25 & 3 & 4 \\ 3 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

$$X^T X - 26I = \begin{bmatrix} -1 & 3 & 4 \\ 3 & -25 & 0 \\ 4 & 0 & -25 \end{bmatrix}$$

→ could find vector in nullsp($X^T X - 26I$)

→ or, could solve

$$\begin{bmatrix} 25 & 3 & 4 \\ 3 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 26a \\ 26b \\ 26c \end{bmatrix}$$

→ $25a + 3b + 4c = 26a$

$\Rightarrow a = 3b + 4c$

$\Rightarrow 3a + b = 26b$

$3a = 25b$

$a = 3b + 4c$

$3a = 25b$

set $c = 1$

$\vec{V}_1 = \begin{bmatrix} 25/4 \\ 3/4 \\ 1 \end{bmatrix} \Rightarrow$ convert to unit vector, becomes first col of V

$a = 3b + 4$

$3(3b+4) = 25b$

$9b + 12 = 25b$

$12 = 16b$

$b = \frac{12}{16} = \frac{3}{4}$

$3a = 25 \cdot \frac{3}{4}$

$a = 25 \cdot \frac{3}{4} \cdot \frac{1}{3}$

$= \frac{25}{4}$

Computing the SVD By Hand

To conclude, we found $X = U\Sigma V^T$ by:

1. Computing X^TX.
2. Finding the eigenvalues of X^TX; their square roots are the singular values of X
.

$$\sigma_i = \sqrt{\lambda_i}$$

   These singular values are placed in the diagonal of Σ in decreasing order.
3. For each eigenvalue λ_i, finding an orthonormal eigenvector $\vec{v}_i$ of X^TX and placing it in the ith column of V.
4. For each $i = 1, 2, ..., r$, finding $\vec{u}_i$ by solving

$$\vec{u}_i = \frac{1}{\sigma_i} X\vec{v}_i$$

   and placing it in the ith column of U.
5. Filling the rest of U with orthonormal vectors that form a basis for $\text{nullsp}(X^T)$.

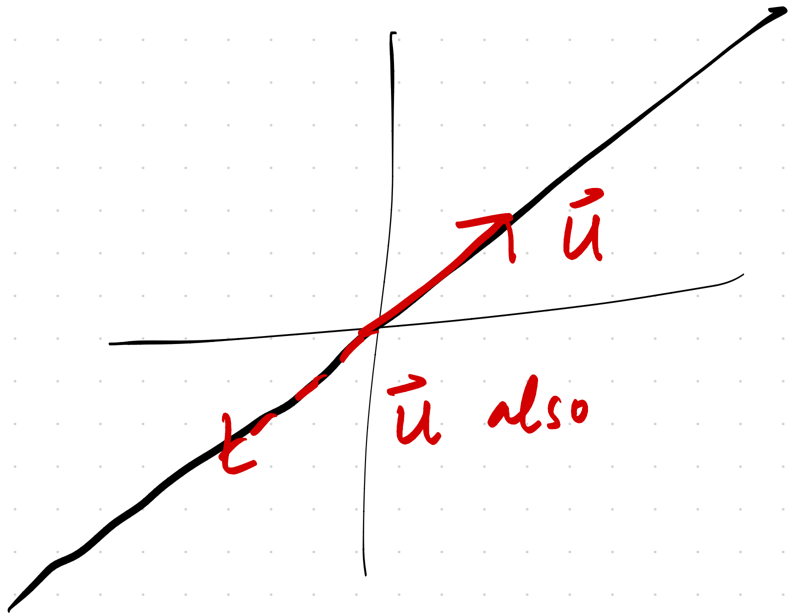
This is not the only possible sequence of steps to follow; for instance, once you find the singular values $\sigma_1, \sigma_2, ..., \sigma_r$, you can independently find orthonormal eigenvectors of X^TX and XX^T and use them to form U and V. Just make sure you place the $\vec{u}_i$'s and $\vec{v}_i$'s in the correct order, corresponding to the order of the singular values in Σ.

σ_i, $\vec{u}_i$, $\vec{v}_i$
triplet of friends, need to satisfy

$$X\vec{v}_i = \sigma_i \vec{u}_i$$

eigvec XX^T

eigvec of X^TX



$\vec{u}$
$\vec{u}$ also

$$
\underbrace{\begin{bmatrix} 3 & 2 & 5 \\ 2 & 3 & 5 \\ 2 & -2 & 0 \\ 5 & 5 & 10 \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{3\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{2}{3} \\ \frac{1}{\sqrt{6}} & -\frac{1}{3\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & 0 & \frac{1}{3} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}}_{V^T}
$$

The summation view of the SVD says that:

$$
\begin{aligned}
X &= \overset{\sigma_1}{15} \overset{\vec{u}_1}{\begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ 0 \\ \frac{2}{\sqrt{6}} \end{bmatrix}} \overset{\vec{v}_1^{\,T}}{\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix}} + 3 \begin{bmatrix} \frac{1}{3\sqrt{2}} \\ -\frac{1}{3\sqrt{2}} \\ \frac{2\sqrt{2}}{3} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} \frac{5}{2} & \frac{5}{2} & 5 \\ \frac{5}{2} & \frac{5}{2} & 5 \\ 0 & 0 & 0 \\ 5 & 5 & 10 \end{bmatrix}}_{\text{rank-one matrix}} + \underbrace{\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 2 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{rank-one matrix}}
\end{aligned}
$$

Frobenius norm : pretend A is a vector and compute its (ℓ_2) norm

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \|A\|_F = \sqrt{1^2+2^2+3^2+4^2}$$

How do we measure how close X_{approx} is from X?

Compute

$$\| X - X_{approx} \|_F$$

SVD says that any rank-r matrix X can be written as a sum of r rank-1 matrices

$$X = \sigma_1 \vec{u}_1 \vec{v}_1^{\,T} + \underbrace{\sigma_2 \vec{u}_2 \vec{v}_2^{\,T}}_{\text{outer product : rank-1}} + \cdots + \sigma_r \vec{u}_r \vec{v}_r^{\,T}$$

$$= \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Low-rank approximation

Suppose X is a rank-r matrix,

$k \leq r$

then the **best** rank-k approximation of X is

$$X_k = \underbrace{\sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T}_{}$$

just take the first k outer products!

X_k minimizes $\|X - X_k\|_F^2$ ← see HW 11 summing up k rank-one matrices.

$$\text{rank } k \text{ approximation of image} = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T$$

A slider should appear below, allowing you to select a value of k and see the corresponding rank-k approximation of Junior.

Rank k: 40

1 7 13 19 25 31 37 43 49

$(1+n+d)k$

Here:

$(1+400+300)40$

$\approx 28\,000$

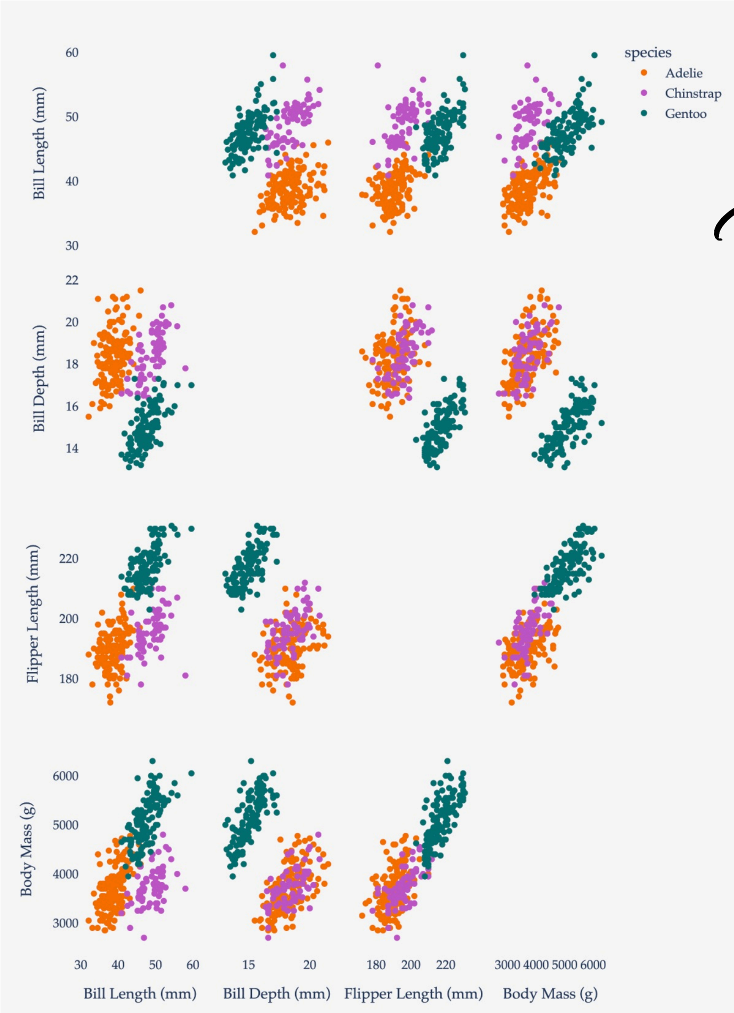
$\ll 120\,000$

(if you stored all 400×300 pixels)

Chapter 10.3

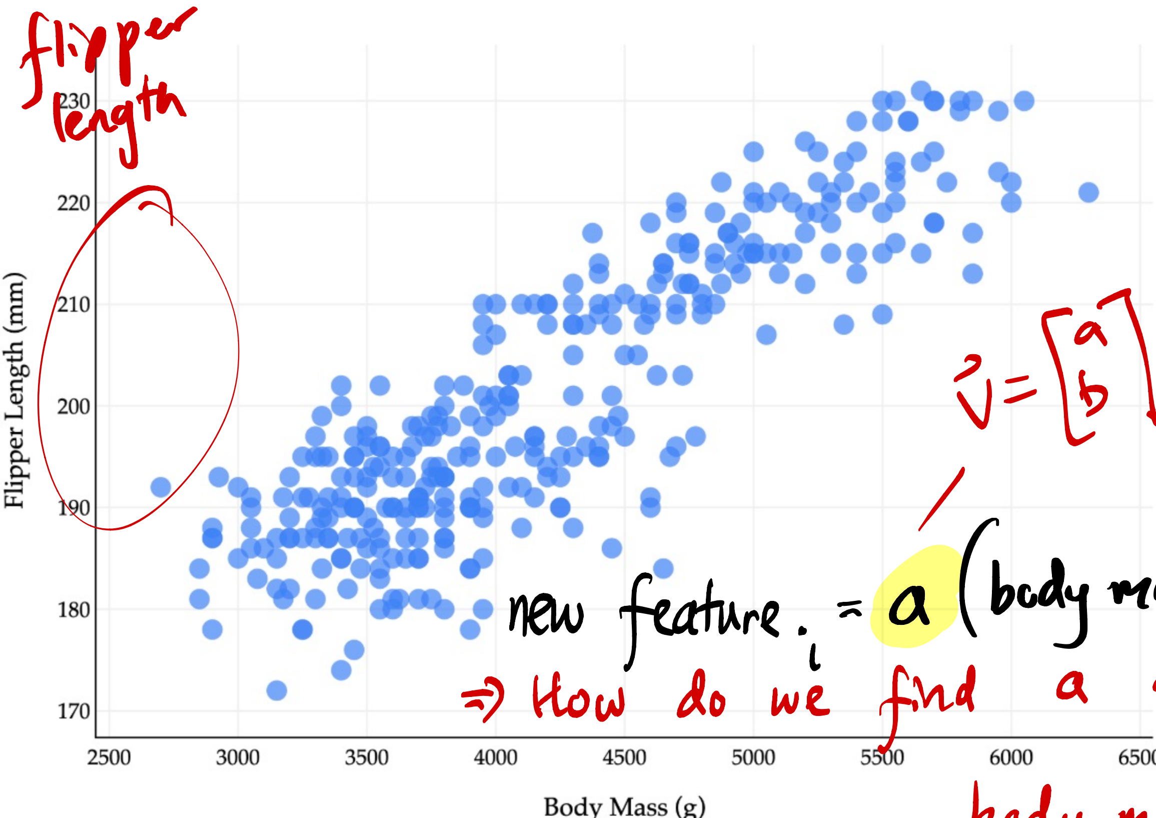
each row : 1 penguin

species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male
Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female
Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female
Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female
Adelie	Torgersen	39.3	20.6	190.0	3650.0	Male
...	...	...	...	...	...	...
Gentoo	Biscoe	47.2	13.7	214.0	4925.0	Female
Gentoo	Biscoe	46.8	14.3	215.0	4850.0	Female
Gentoo	Biscoe	50.4	15.7	222.0	5750.0	Male
Gentoo	Biscoe	45.2	14.8	212.0	5200.0	Female
Gentoo	Biscoe	49.9	16.1	213.0	5400.0	Male



hard to deal with several numerical features simultaneously

- idea: what if we create one new feature using all of the existing features?



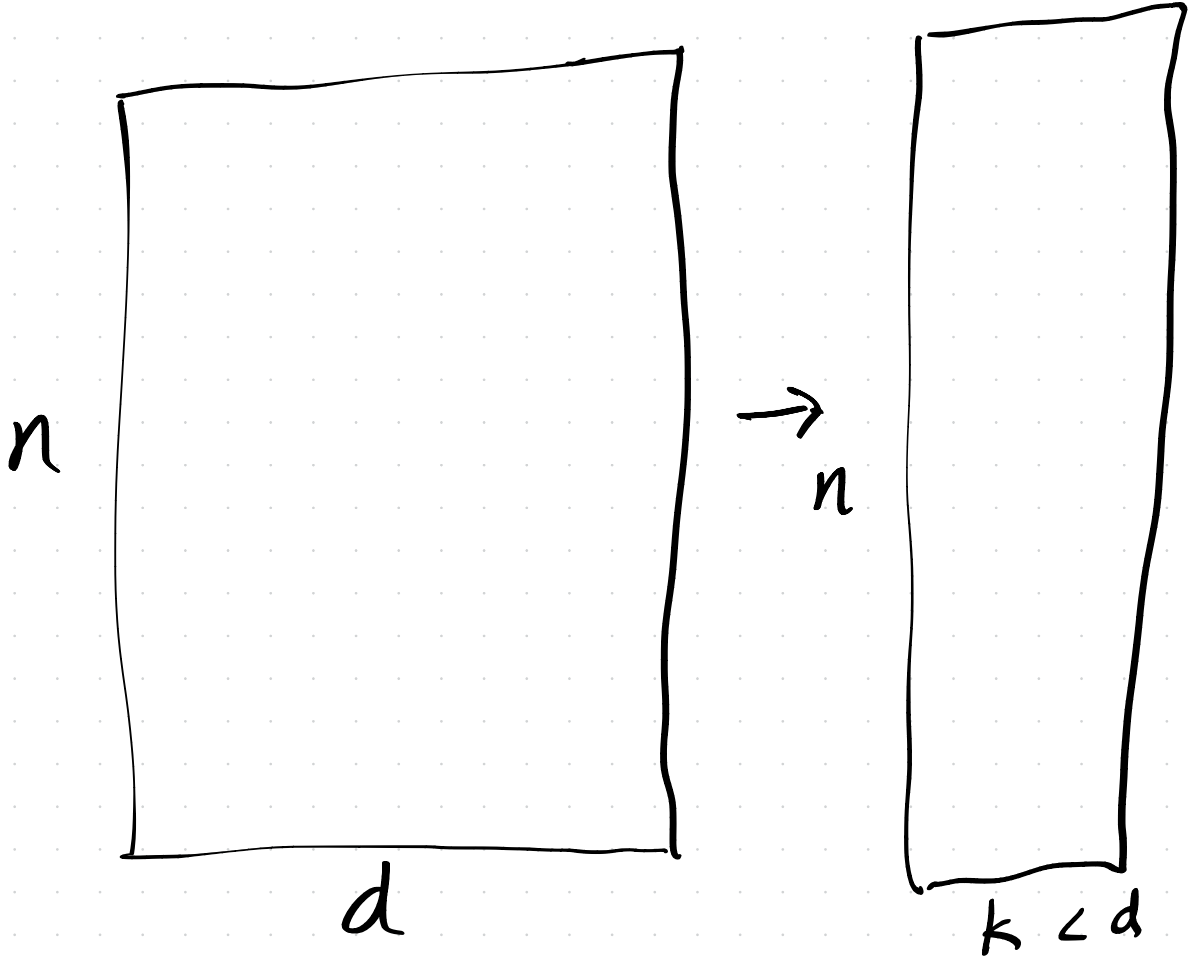
Goal: produce one new feature using a combination of both existing features

$$\vec{v} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\text{new feature}_i = a(\text{body mass}_i) + b(\text{flipper length}_i)$$

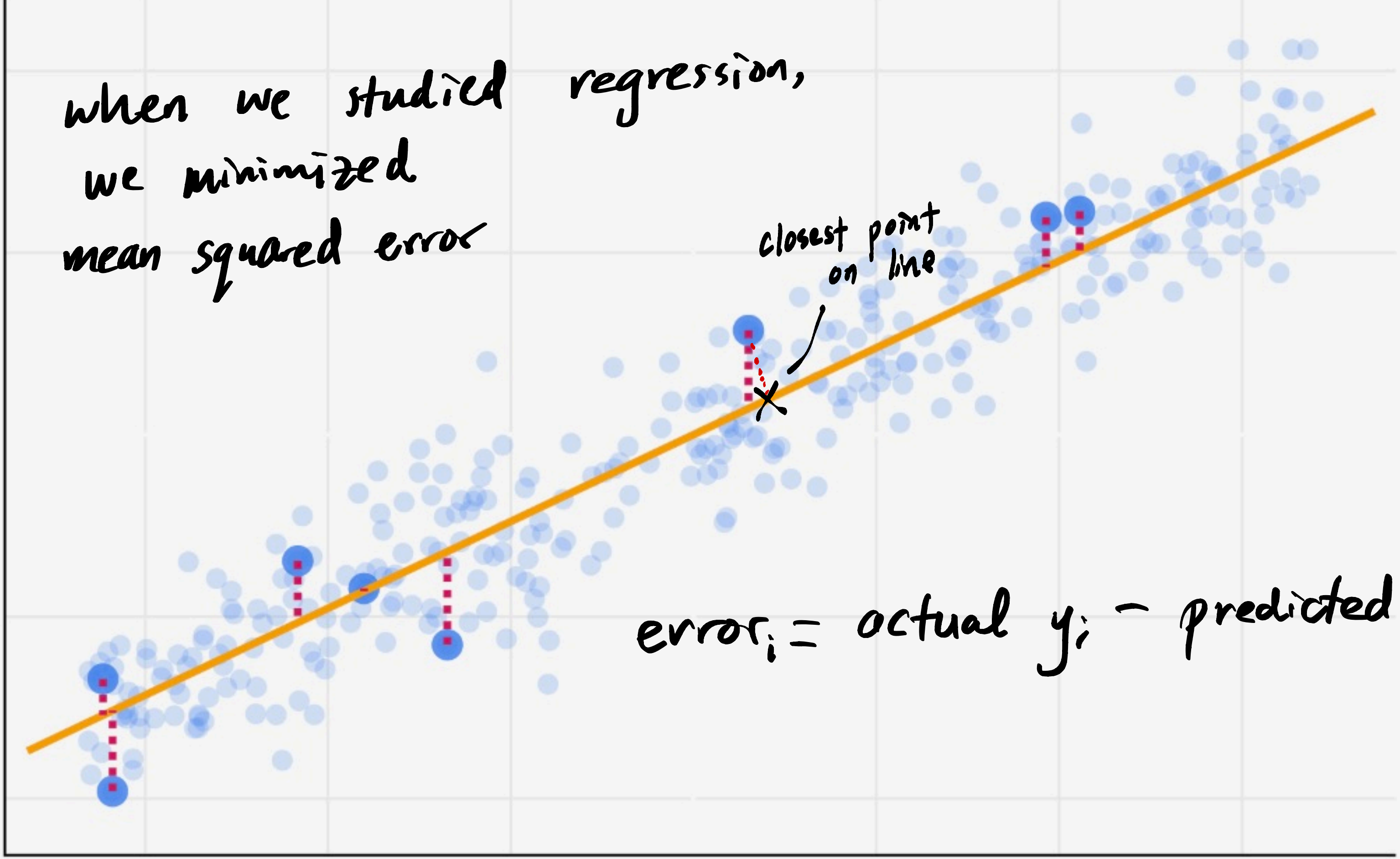
⇒ How do we find a and b?

body mass

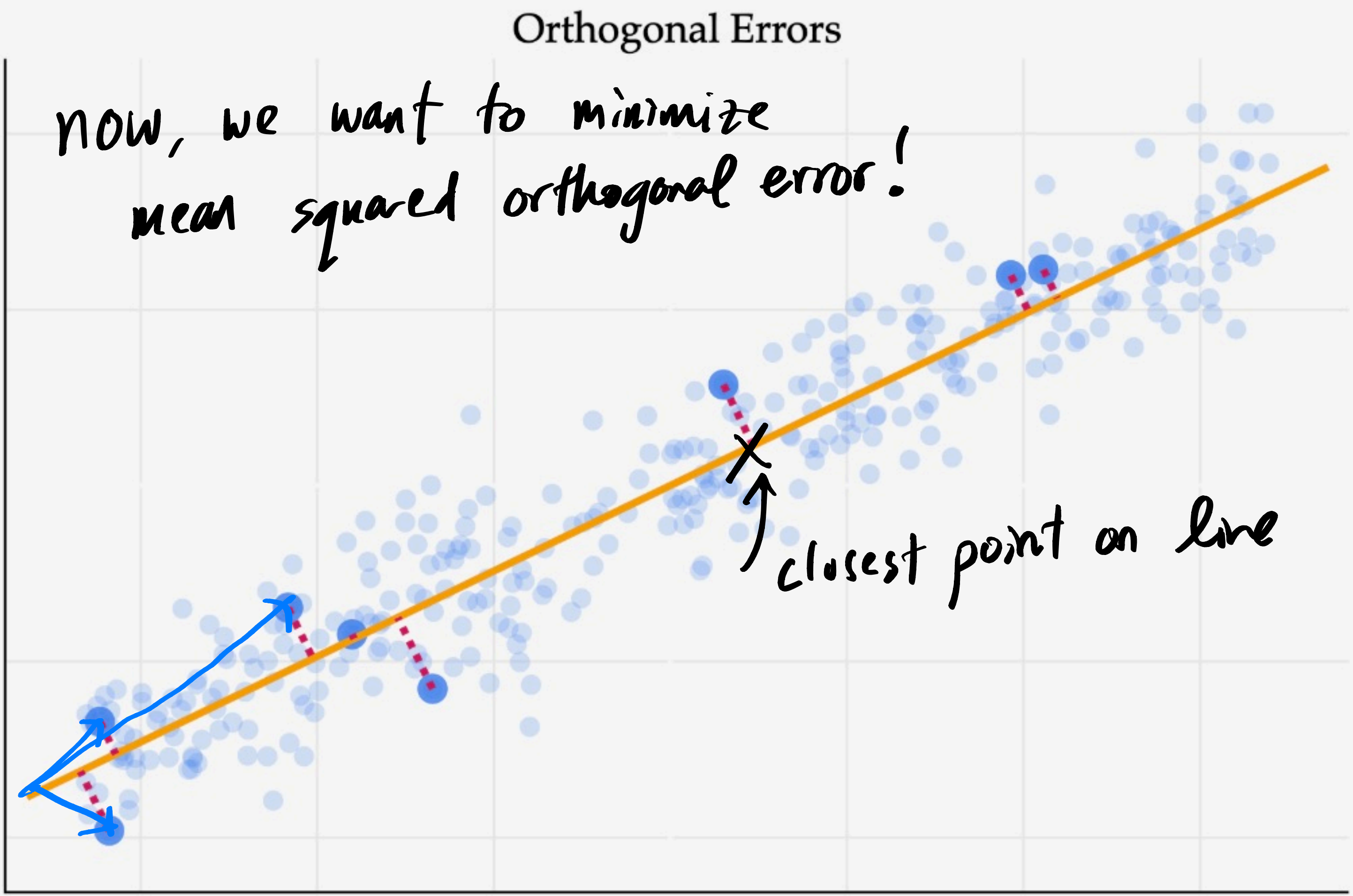


this is dimensionality reduction

which is a form of **unsupervised learning** when we studied regression, we minimized mean squared error



$$\text{error}_i = \text{actual } y_i - \text{predicted } y_i$$


Orthogonal Errors
now, we want to minimize
mean squared orthogonal error!
closest point on line

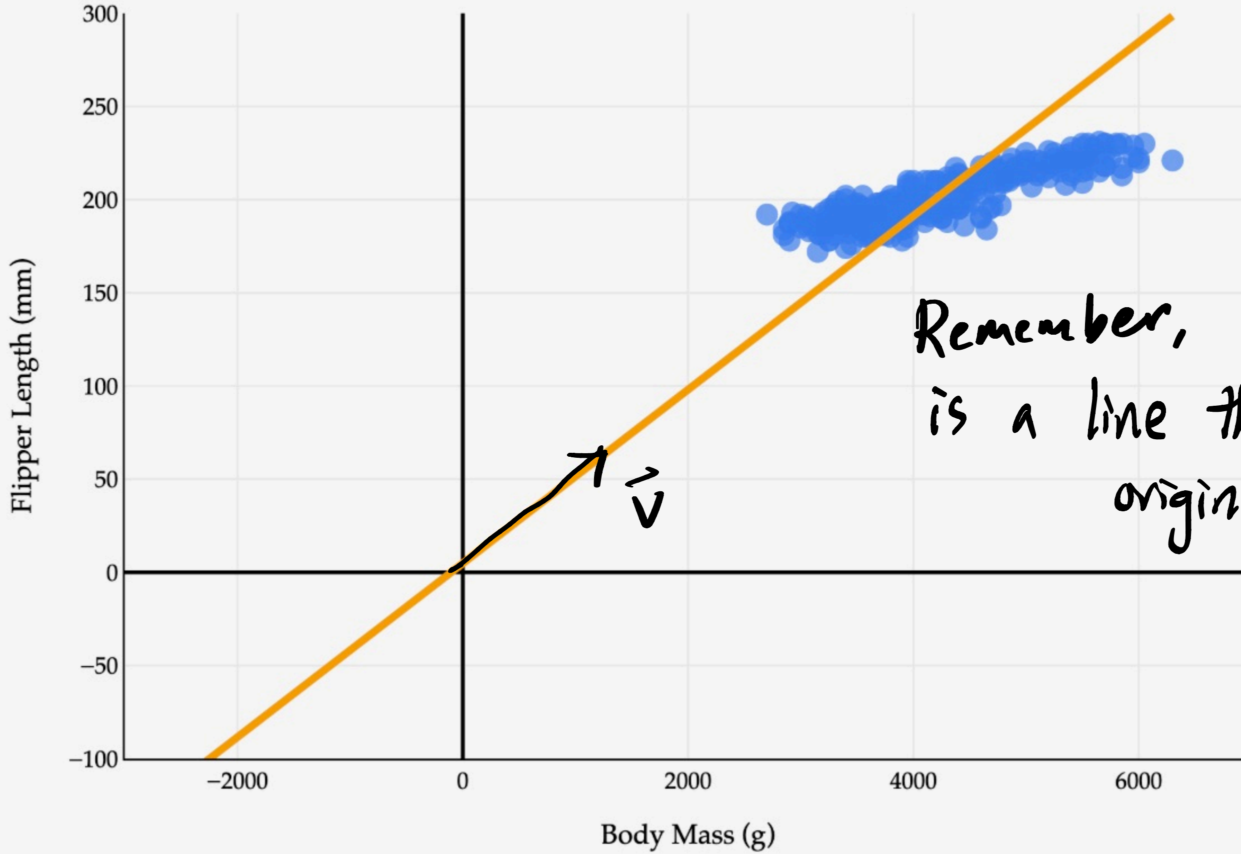
Key idea : we have hundreds of vectors (rows of dataset),

all being projected onto one vector : $\vec{v}$.

Q: what is the best $\vec{v}$?

A: $\vec{v}$ that minimizes mean squared orthogonal error!

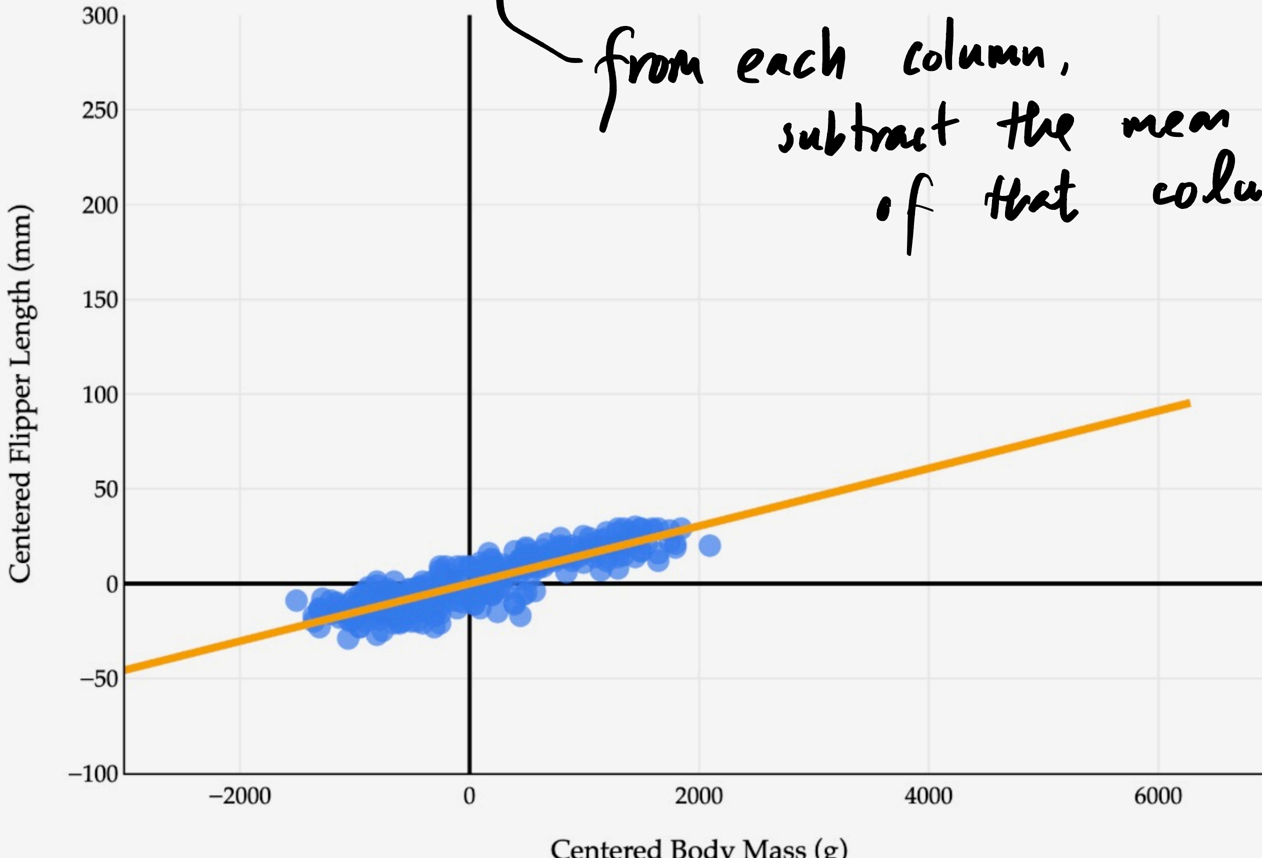
Our data isn't usually located near the origin...



Remember, span($\{\vec{v}\}$) is a line through the origin!

...which is why we center the data first! This doesn't change its shape.

from each column, subtract the mean of that column

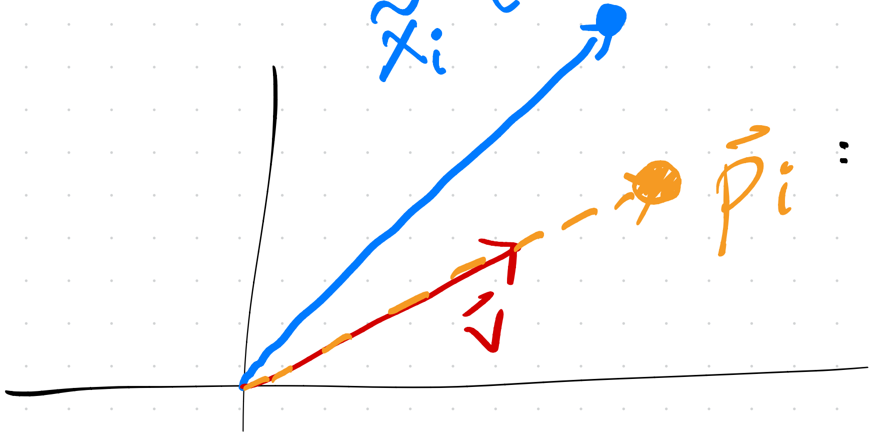


Suppose X is our original $n \times d$ matrix.

$\tilde{X}$ is the mean-centered version of X, produced by subtracting the mean of each column from each column

What is $\tilde{X}^T \vec{1}$? $\tilde{X}^T \vec{1}$ contains the sum of each col of $\tilde{X}$, each sum is 0, so $\tilde{X}^T \vec{1} = \vec{0}$!

Remember, goal is to find the best $\vec{v}$!



$\tilde{x}_i$ (vector, row i of $\tilde{X}$)

$\vec{p}_i$: projection of $\tilde{x}_i$ onto $\vec{v}$

$$\vec{p}_i = \left(\frac{\tilde{x}_i \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \right) \vec{v}$$

For simplicity, since its length doesn't matter, assume $\|\vec{v}\| = 1$.

Then,

$$\vec{p}_i = (\tilde{x}_i \cdot \vec{v}) \, \vec{v}$$

$$\vec{p}_i = (\tilde{x}_i \cdot \vec{v})\,\vec{v}$$

Goal: Minimize mean squared orthogonal error:

Formula?

$$\frac{1}{n}\sum_{i=1}^{n} \left\| \tilde{x}_i - \vec{p}_i \right\|^2$$

$$J(\vec{v}) = \frac{1}{n}\sum_{i=1}^{n} \left\| \tilde{x}_i - (\tilde{x}_i \cdot \vec{v})\,\vec{v} \right\|^2$$

"objective function"   what $\vec{v}$ minimizes $J(\vec{v})$?

Observation: if $\vec{v}$ minimizes mean squared projection error, it maximizes the variance of the $\vec{p}_i$'s!

Why is that?

$$J(\vec{v}) = \frac{1}{n}\sum_{i=1}^{n} \| \tilde{x}_i - (\tilde{x}_i \cdot \vec{v})\vec{v} \|^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} \left(\tilde{x}_i - (\tilde{x}_i \cdot \vec{v})\vec{v} \right) \cdot \left(\tilde{x}_i - (\tilde{x}_i \cdot \vec{v})\vec{v} \right)$$

$$= \vdots \quad \text{algebra skipped}$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n} \|\tilde{x}_i\|^2}_{\text{constant!}} - \frac{1}{n}\sum_{i=1}^{n} (\tilde{x}_i \cdot \vec{v})^2$$

So, minimizing $J(\vec{v}) = \frac{1}{n}\sum_{i=1}^{n} \|\tilde{x}_i - (\tilde{x}_i \cdot \vec{v})\vec{v}\|^2$

is the same as maximizing

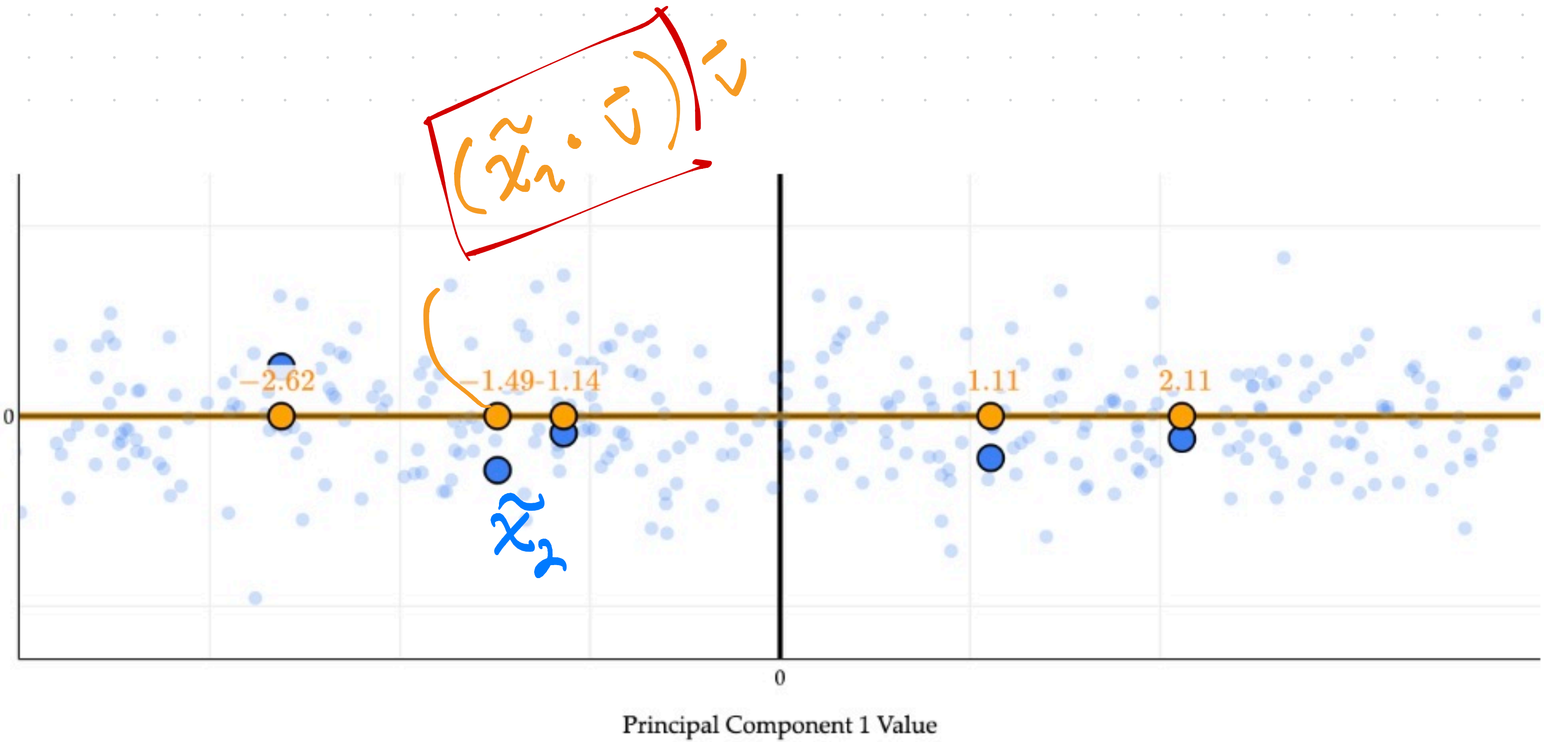
$$PV(\vec{v}) = \frac{1}{n}\sum_{i=1}^{n} (\tilde{x}_i \cdot \vec{v})^2$$

"projected variance" of new feature values

$$\text{new feature}_i = \tilde{x}_i \cdot \vec{v}$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

but these new feature values have a mean of 0!



$\boxed{(\tilde{x}_2 \cdot \vec{v})}\vec{v}$
$\tilde{x}_2$
−2.62
−1.49
-1.14
1.11
2.11
0
0
Principal Component 1 Value

$$PV(\vec{v}) = \frac{1}{n}\sum_{i=1}^{n}(\tilde{x}_i \cdot \vec{v})^2$$

projected variance

→ Goal: Find the $\vec{v}$ that maximizes $PV(\vec{v})$

→ <u>why</u> is this easier than finding the $\vec{v}$ that minimizes $J(\vec{v})$?

→ Insight:

$$PV(\vec{v}) = \frac{1}{n}\sum_{i=1}^{n}(\tilde{x}_i \cdot \vec{v})^2 = \frac{1}{n}\|\tilde{X}\vec{v}\|^2$$

→ How does this help?

$$\tilde{X} = \begin{bmatrix} - \tilde{x}_1^T - \\ - \tilde{x}_2^T - \\ \vdots \end{bmatrix}$$

Aside

$\tilde{x}_1 \cdot \vec{v}$, $\tilde{x}_2 \cdot \vec{v}$, ---, $\tilde{x}_n \cdot \vec{v}$

→ These are the dot products of the rows of $\tilde{X}$ with $\vec{v}$

→ These all get computed when we multiply

$$\tilde{X}\vec{v} = \begin{bmatrix} \tilde{x}_1 \cdot \vec{v} \\ \tilde{x}_2 \cdot \vec{v} \\ \vdots \\ \tilde{x}_n \cdot \vec{v} \end{bmatrix}$$

Goal: Maximize

$$PV(\vec{v}) = \cancel{\frac{1}{n}} \|\tilde{X}\vec{v}\|^2,$$

or equivalently just $\|\tilde{X}\vec{v}\|^2$

Subject to the constraint that $\|\vec{v}\| = 1$

→ Issue: constrained optimization is complicated, not in scope for us

→ but ... there's a solution!

Goal:

maximize $\|\tilde{X}\vec{v}\|^2$ subject to $\|\vec{v}\| = 1$

Equivalent goal:

maximize $f(\vec{v}) = \dfrac{\|\tilde{X}\vec{v}\|^2}{\|\vec{v}\|^2}$

if $\vec{v}$ not a unit vector, its length cancels out in $\dfrac{\text{numerator}}{\text{denominator}}$

Goal: maximize

$\vec{v} \neq \vec{0}$

$$f(\vec{v}) = \frac{\|\tilde{X}\vec{v}\|^2}{\|\vec{v}\|^2} = \frac{\vec{v}^T \tilde{X}^T \tilde{X} \vec{v}}{\|\vec{v}\|^2}$$

see HW 8 gradient problem for derivation

Now, take gradient of $f(\vec{v})$!

$$\nabla f(\vec{v}) = \frac{2}{\|\vec{v}\|^2}\left(\tilde{X}^T X \vec{v} - f(\vec{v})\vec{v}\right)$$

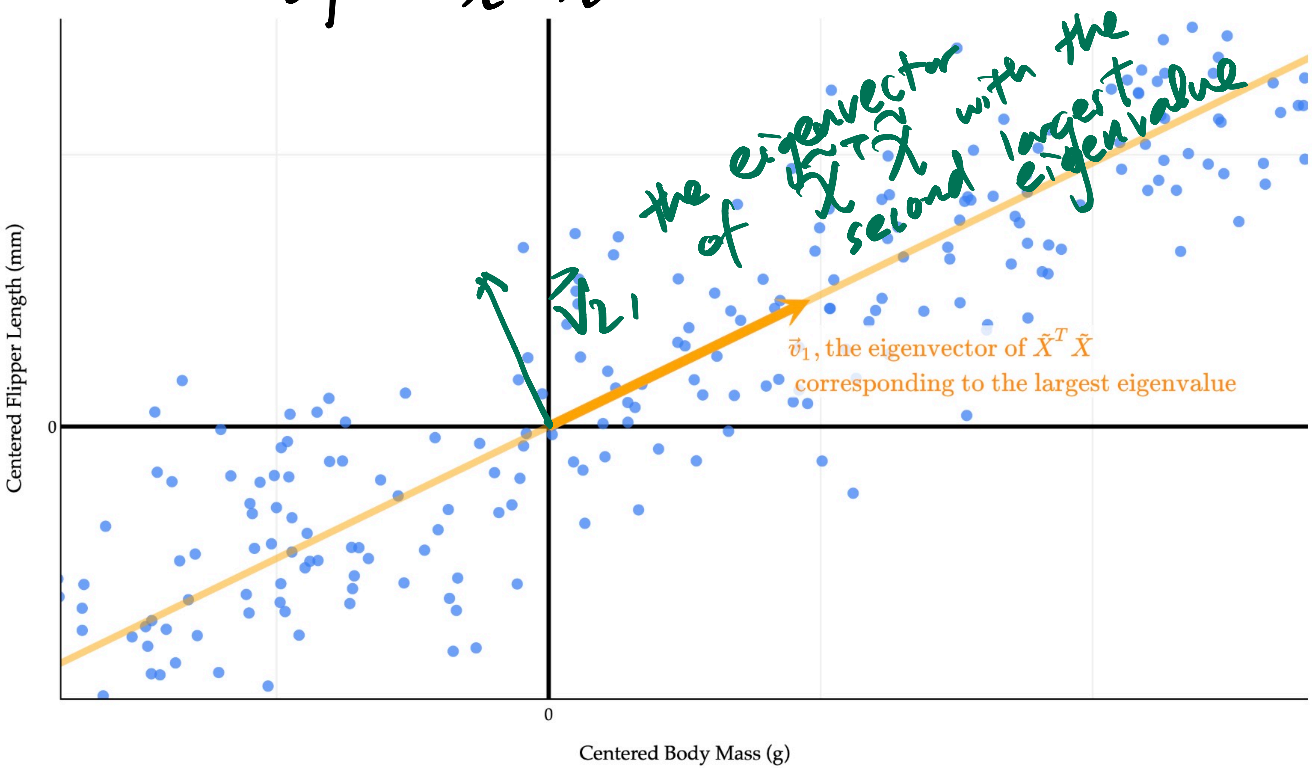
set to 0: $\tilde{X}^T X \vec{v} - f(\vec{v})\vec{v} = \vec{0}$

$$\tilde{X}^T \tilde{X} \vec{v} = f(\vec{v})\vec{v}$$

$$\tilde{X}^T \tilde{X} \vec{v} = f(\vec{v}) \vec{v}$$

- If $\vec{v}$ maximizes $f(\vec{v})$, it satisfies the above equation
- Look carefully: it is saying $\vec{v}$ is an <u>eigenvector of $\tilde{X}^T \tilde{X}$!</u>.
- If we want to maximize $f(\vec{v})$, choose $\vec{v}$ to be the eigenvector with the <u>largest eigenvalue!</u>.

Important: in $\tilde{X} = U\Sigma V^T$ (the SVD of $\tilde{X}$), the first column of V is the eigenvector of $\tilde{X}^T\tilde{X}$ with the largest eigenvalue!!!



$$(\text{new feature } 1)_i = (\tilde{X}\vec{v}_1)_i = \tilde{x}_i \cdot \vec{v}_1$$

where $\vec{v}_1$ is first column of V in

$$\tilde{X} = U \Sigma V^T$$

$V^T V = I$, so the columns of V are unit vectors and orthogonal to one another

$$(\text{new feature } 2)_i = (\tilde{X}\vec{v}_2)_i$$

Principal components analysis

- "Principal component" = "new feature"
- "Principal components analysis" = act of producing several new features, e.g. for visualization, or for model building

To do PCA, first, construct $\tilde{X}$ by mean-centering X.

Then, $\tilde{X} = U \Sigma V^T$ find the SVD of $\tilde{X}$.

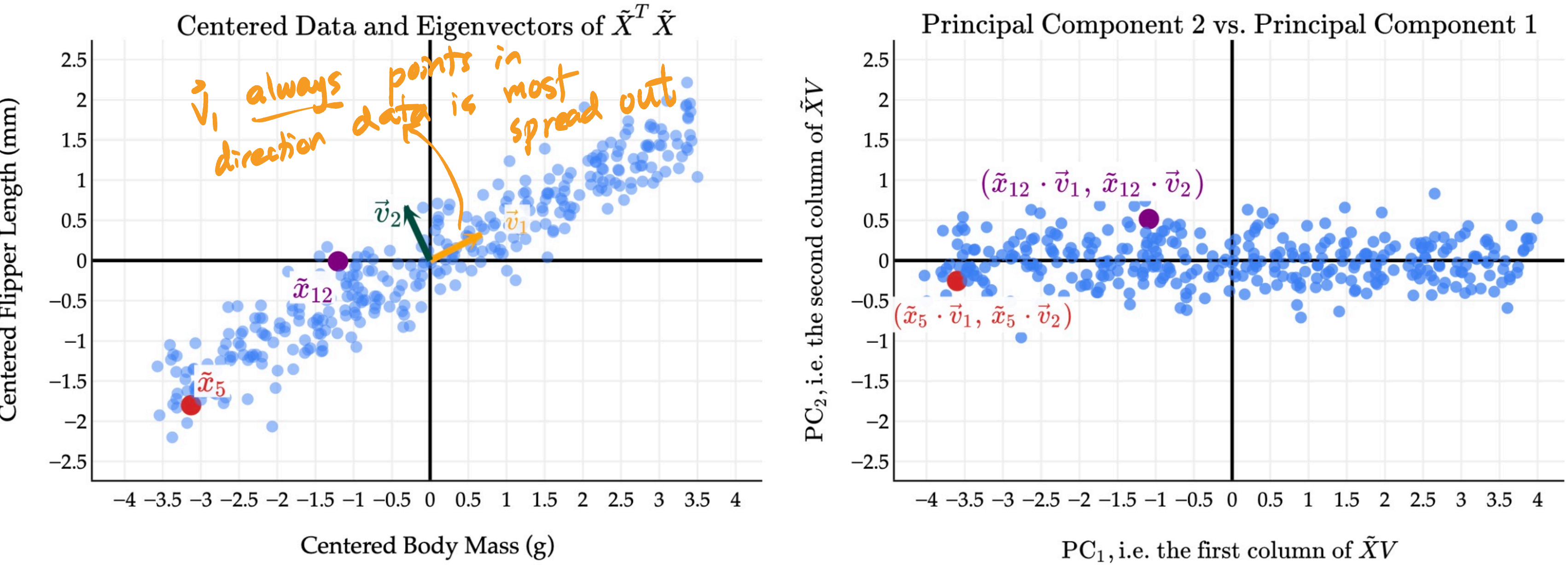
Then,

$$PC_j = \tilde{X} \vec{v}_j = \sigma_j \vec{u}_j$$

(new feature j)

my data

tells me what linear combination of my original features is "best"


Centered Data and Eigenvectors of $\tilde{X}^T \tilde{X}$
$\vec{v}_1$ always points in direction data is most spread out
$\vec{v}_2$
$\vec{v}_1$
$\tilde{x}_{12}$
$\tilde{x}_5$
Centered Flipper Length (mm)
Centered Body Mass (g)
Principal Component 2 vs. Principal Component 1
$(\tilde{x}_{12} \cdot \vec{v}_1, \tilde{x}_{12} \cdot \vec{v}_2)$
$(\tilde{x}_5 \cdot \vec{v}_1, \tilde{x}_5 \cdot \vec{v}_2)$
PC_2, i.e. the second column of $\tilde{X}V$
PC_1, i.e. the first column of $\tilde{X}V$