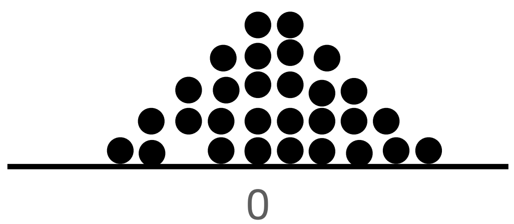
Notation and Math

Gaussian Distribution (bell curve)

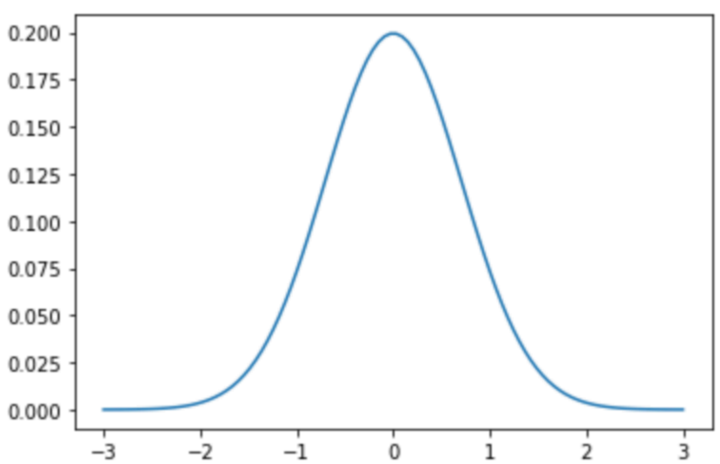
Intuition

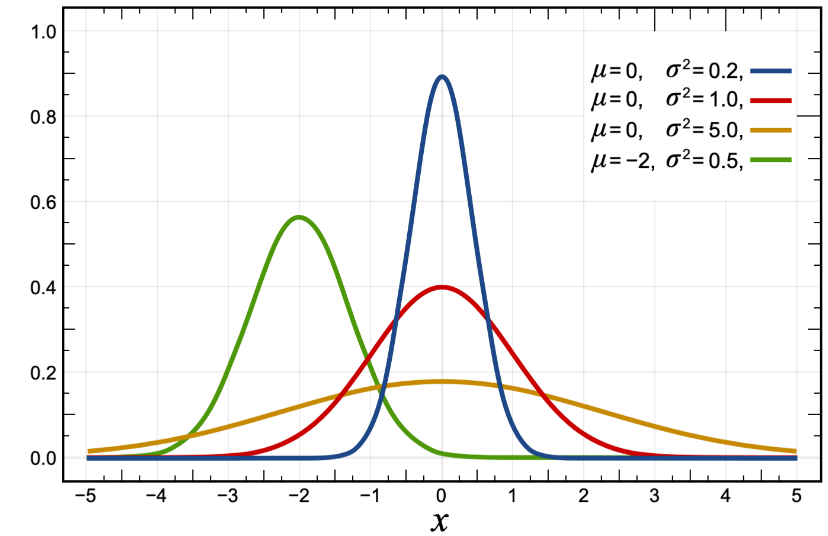
$y^{(i)} = \mathbf{w}^T \mathbf{x}^{(i)} + \epsilon^{(i)}$

We need this next class



$\epsilon^{(i)}$ Is the residual





https://commons.wikimedia.org/wiki/File:Normal_Distribution_PDF.svg

$$\frac{1}{\sqrt{2\pi}} e^{-(\epsilon^{(i)})^2/2}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(k-\mu)^2/2\sigma^2}$$

$\epsilon^{(i)}$ drawn from $\mathcal{N}(0,1)$

We need this next class

Properties of the log function

We will use these in some of the lectures

- $\log abc = \log a + \log b + \log c$
- $\log \prod_{i=1}^{3} \epsilon^{(i)} = \log(\epsilon^{(1)}\epsilon^{(2)}\epsilon^{(3)}) = \log \epsilon^{(1)} + \log \epsilon^{(2)} + \log \epsilon^{(3)} = \sum_{i=1}^{3} \log \epsilon^{(i)}$
- $\log a^c = c \log a$

Matrix Inverse

- Any nonzero number has an inverse: e.g. 3 has an inverse 1/3
- Square matrices with linearly independent columns (or independent rows) have inverses. E.g.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} \text{ since}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$X^T X$ is a square matrix

If the columns (features) are independent $X^T X$ has an inverse

- If X is a $N \times (d+1)$ matrix then:
  - X^T is a $(d+1) \times N$ matrix
  - $X^T X$ is a $(d+1) \times (d+1)$ matrix

Matrix Pseudo Inverse

Generalization of the inverse, defined for all matrices

- Some matrices don't have inverses. For example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

However it has a "left inverse" $\begin{bmatrix} -4/3 & -1/3 & 2/3 \\ 13/12 & 1/3 & -5/12 \end{bmatrix}$

$$\begin{bmatrix} -4/3 & -1/3 & 2/3 \\ 13/12 & 1/3 & -5/12 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- But.... $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -4/3 & -1/3 & 2/3 \\ 13/12 & 1/3 & -5/12 \end{bmatrix} = \begin{pmatrix} \frac{5}{6} & \frac{1}{3} & -\frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{6} & \frac{1}{3} & \frac{5}{6} \end{pmatrix}$

Pseudo Inverse continued

Let X be a $N \times (d+1)$ matrix with independent columns

- $X^T X$ is a square matrix (X^T is a $(d+1) \times N$ matrix - thus $X^T X$ is a $(d+1) \times (d+1)$ matrix)
- The pseudo-inverse (Moore-Penrose inverse) exists and is unique: $X^+ = (X^T X)^{-1} X^T$
- For linear regression, the goal is to find $\mathbf{w}$ such that $\mathbf{y} \approx X\mathbf{w}$. If X is a $N \times (d+1)$ matrix then
  - setting $\mathbf{w} = X^+ \mathbf{y}$ produces the smallest $\|X\mathbf{w} - \mathbf{y}\|_2 = \|\hat{\mathbf{y}} - \mathbf{y}\|_2$
  - If we don't have independent columns - we still find $\mathbf{w}$ that minimizes $\|X\mathbf{w} - \mathbf{y}\|_2$ (but it is not unique)

np.linalg.pinv is the numpy function to find the pseudo inverse

https://inst.eecs.berkeley.edu/~ee127/sp21/livebook/def_pseudo_inv.html

https://people.cs.umass.edu/~cmusco/personal_site/pdfs/linear_least_squares.pdf

I added the next slide if you wondered about how you could calculate the pseudo inverse (left inverse) using the SVD.

You are not responsible for the following material

Calculating the Pseudo Inverse

Let A be a $n \times m$ matrix

- If the columns of A are independent (i.e. rank(A) = m) then A has a *left inverse:*

$$A^+ = (A^T A)^{-1} A^T.$$

- Thus $A^+ A = I$ where I is an $m \times m$ identity matrix.

- FYI, we can calculate the pseudo inverse by first calculating the SVD of $A = U \begin{bmatrix} S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} V^T$ where U, V are orthogonal matrices and S is a diagonal matrix (all values on the diagonal are positive) and then compute:

$$A^+ = (A^T A)^{-1} A^T = V \begin{bmatrix} S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} U^T$$