EE2211 Tutorial 4

$X: \mathbb{R}^d \to \mathbb{R}^m$

$d \sim$ dim. feature; samples $\sim m$; $\underbrace{X}\,\underbrace{w} = y$

(Systems of Linear Equations)
Question 1:

$X: \mathbb{R}^2 \to \mathbb{R}^2$

Given $\mathbf{X}\mathbf{w} = \mathbf{y}$ where $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(a) What kind of system is this? (even-, over- or under-determined?)
(b) Is $\mathbf{X}$ invertible? Why? yes. $\det X \neq 0$
(c) Solve for $\mathbf{w}$ if it is solvable. $\hat{w} = X^{-1}y$.

$rk\,X = 2 = $ full. Row Ech Form $\sim \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

(Systems of Linear Equations)
Question 2:

$m = 2,\ d = 2.$

Given $\mathbf{X}\mathbf{w} = \mathbf{y}$ where $\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

$\lambda \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

(a) What kind of system is this? (even-, over- or under-determined?)
(b) Is $\mathbf{X}$ invertible? Why? No.
(c) Solve for $\mathbf{w}$ if it is solvable.

① LI exist? No. Not inj.
② RI exist? No. Not surj.

$\det(X^TX) = 0$
$\det(XX^T) = 0$

(Systems of Linear Equations)
Question 3:

$X: \mathbb{R}^2 \to \mathbb{R}^3$

$Xw = y$

$3\overset{2}{\begin{pmatrix} x_{11} & x_{12} \\ & \end{pmatrix}}\, 2\overset{1}{\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}} = 3\overset{1}{\begin{pmatrix} y_1 \\ \vdots \\ \vdots \end{pmatrix}}$

Given $\mathbf{X}\mathbf{w} = \mathbf{y}$ where $\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 1 & -1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 0 \\ 0.1 \\ 1 \end{bmatrix}$.

no. of samples $\overbrace{m = 3}$
$d = 2$ no. of features
$m > d$.

(a) What kind of system is this? (even-, over- or under-determined?)
(b) Is $\mathbf{X}$ invertible? Why? No. $m \neq d$. not full rk.
(c) Solve for $\mathbf{w}$ if it is solvable.

Left Inverse.

① check X^TX invertible? Calc. $\det(X^TX) \neq 0$. $\Rightarrow \therefore X^TX$ is invertible; LI exist.
② Set $\hat{w} = (X^TX)^{-1}X^Ty$.

$X^TX = \begin{bmatrix} 6 & 9 \\ 9 & 21 \end{bmatrix}$

(Systems of Linear Equations)
Question 4:

$XX^T = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

$X: \mathbb{R}^d \to \mathbb{R}^m$

Given $\mathbf{X}\mathbf{w} = \mathbf{y}$ where $\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

$m = 3$
$d = 4$

(a) What kind of system is this? (even-, over- or under-determined?)
(b) Is $\mathbf{X}$ invertible? Why? No. $m \neq d$.
(c) Solve for $\mathbf{w}$ if it is solvable. Apply RI. $\hat{w} = X^T(XX^T)^{-1}y$

(Systems of Linear Equations)
Question 5:

Invertibility = X^{-1} exists

How to check invertibility

① * (know this) X is invertible $\Leftrightarrow \frac{1}{\det X}$ exist (we code) $\Leftrightarrow \det X \neq 0$

② Invertible $\Leftrightarrow$ Bijective (one-to-one & onto)
on linear maps: Bijective requires $\boxed{m = d}$ *

②b How to check 1-1 & onto?
extra: 1-1: column vectors need to be linearly independent
onto: column vectors need to be spanning
i.e. for every $v \in \mathbb{R}^m$, we can write $v = c_1X_1 + \dots + c_nX_n$
where $X = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}$
$\therefore$ Invertible $\Leftrightarrow$ columns are basis vectors for codomain

③ Invertible $\Leftrightarrow$ has left-inverse & right-inverse
(Extra) can use row-echelon form
$X \sim \begin{pmatrix} 1 & & * \\ 0 & 1 & \\ & & \ddots \end{pmatrix}$

1-1: Has Left-inverse: pivots in every column. e.g. $\begin{pmatrix} 1 & 7 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ vs negative e.g. $\begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$

onto: Has Right-inverse: pivots in every row e.g. $\begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix}$ vs neg. e.g. $\begin{pmatrix} 1 & 3 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$

Invertible = pivots in all rows & columns

What is the left/right inverse?

$\mathbb{R}^d \xrightarrow{X} \mathbb{R}^m \xrightarrow{X^L} \mathbb{R}^d$ id.

Left-inverse: X^L satisfies $X^LXw = w$.
* Formula: $X^L = (\underset{\sim}{X^TX})^{-1}X^T$ → is this correct? Gut check.
* Left-inverse exist $\Leftrightarrow$ X^TX is invertible $\Leftrightarrow \det X^TX \neq 0$

Right-Inverse: X^R satisfies $XX^Ry = y$
* Formula: $X^R = X^T(XX^T)^{-1}$
* Right-inverse exist $\Leftrightarrow$ $\underline{XX^T \text{ invertible}} \Leftrightarrow \underline{\det XX^T \neq 0}$

$m > d$.

$X^LX = I$.

$\left[(X^TX)^{-1}X^T\right]X$

$= (X^TX)^{-1}X^TX = I$.

dxd matrix.

$X^T w = y$.

set $A = X^T$.

$A = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$

Given $\mathbf{w}^T\mathbf{X} = \mathbf{y}^T$ where $\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(a) What kind of system is this? (even-, over- or under-determined?)
(b) Is $\mathbf{X}$ invertible? Why?
(c) Solve for $\mathbf{w}$ if it is solvable.

take transpose on both sides

$$\left.\begin{aligned} w^T X &= y^T \\ (w^T X)^T &= (y^T)^T \end{aligned}\right\} \quad X^T w = y$$

Set $A = X^T$.

(Systems of Linear Equations)

Question 6: $(w_1, w_2)\begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 1 & -1 \end{pmatrix} = (0 \;\; 1)$

m = no. of samples = 2.

d = no. of features = 3.

Given $\mathbf{w}^T\mathbf{X} = \mathbf{y}^T$ where

$A: \mathbb{R}^3 \to \mathbb{R}^2$.

$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 1 & -1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. $(\cdot)^T X: \mathbb{R}^{d=3} \to \mathbb{R}^{m\,2}$

$A = (\;\;3\;\;)\;(\;)\;=(\;)$

"RI": $X^T(XX^T)^{-1}$

(a) What kind of system is this? (even-, over- or under-determined?)
(b) Is $\mathbf{X}$ invertible? Why? No.
(c) Solve for $\mathbf{w}$ if it is solvable.

$$\begin{aligned} w &= A^T(AA^T)^{-1}y \\ &= (X^T)^T(X^T(X^T)^T)^{-1}y \\ &= X(X^TX)^{-1}y. \end{aligned}$$

(Systems of Linear Equations)

Question 7:

This question is related to determination of types of system where an appropriate solution can be found subsequently. The following matrix has a left inverse.

$\mathbf{X} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(Method 1) $\det(X^TX) = 0$.

(Method 2) NOT injective.

but has Right Inverse.

$X\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

a) True
b) False

What is the left/right inverse?

Left-inverse: X^L satisfies $X^L X w = w$. $\mathbb{R}^d \xrightarrow{X} \mathbb{R}^m \xrightarrow{X^L} \mathbb{R}^d$ (i.d.)

* Formula: $X^L = (X^TX)^{-1}X^T$ → is this correct? Gut check.

* Left-inverse exist $\Leftrightarrow$ X^TX is invertible $\Leftrightarrow$ $\det X^TX \neq 0$

Right-Inverse: X^R satisfies $XX^R y = y$

* Formula: $X^R = X^T(XX^T)^{-1}$

Right-inverse exist $\Leftrightarrow$ ________ $\Leftrightarrow$ ________

(Systems of Linear Equations)

Question 8:

MCQ: Which of the following is/are true about matrix $\mathbf{A}$ below? **There could be more than one answer.**

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

1-1 = Has Left-inverse: pivots in every column. e.g. $\begin{pmatrix} 1 & 7 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ vs $\begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$ negative e.g.

a) $\mathbf{A}$ is invertible No. $m \neq d$.
b) $\mathbf{A}$ is left invertible $\det(X^TX) = 0$. No.
c) $\mathbf{A}$ is right invertible $\det(XX^T) \neq 0$. yes.
d) $\mathbf{A}$ has no determinant Yes. $m \neq d$.
e) None of the above

$A \sim \begin{pmatrix} 1 & * & * \\ 0 & 1 & * \end{pmatrix}$

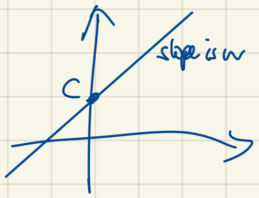
$$A\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\begin{pmatrix} 1 \\ 4 \end{pmatrix} + y\begin{pmatrix} 2 \\ 5 \end{pmatrix} + z\begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

$$\left.\begin{aligned} A\begin{pmatrix} \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{pmatrix} &= \frac{1}{2}\begin{pmatrix} 1 \\ 4 \end{pmatrix} - \begin{pmatrix} 2 \\ 5 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 3 \\ 6 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 4 \\ 10 \end{pmatrix} - \begin{pmatrix} 2 \\ 5 \end{pmatrix} = 0. \end{aligned}\right\} \quad \begin{pmatrix} \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{pmatrix} \neq 0. \text{ Not L.I.}$$

"Machine Learning" algo to find best f

input

$$X \longrightarrow \boxed{} \longrightarrow y_{pred}$$



① Impose structural assumptions on model:

e.g. (linear regression) f is a linear fn, say $y \approx f(x) = wx + c$ (weight) or $y \approx f(x) = wx$

e.g. (polynomial regression) f is a polynomial fn, e.g. $y = f(x) = w_2 x^2 + w_1 x + c$.

simplest. Today, we do this.

Available data: X_train, y_train

② objective : find good estimate for w & c.

a.k.a. : solve for X_train $\cdot$ w = y_train.

here: view X as a linear function $X : \mathbb{R}^d \to \mathbb{R}^m$. we want to solve for best w in $\mathbb{R}^d$.

Note: $y = x \cdot w + c$ is equivalent to $y = \underbrace{(1 \;\; x) \cdot \begin{pmatrix} c \\ w \end{pmatrix}}_{1 \cdot c + x \cdot w}$

③ Available data : X_test.

- use our estimate of w to get

y_pred = X_test $\cdot\, w$

Here, view $w : \mathbb{R}^{m \times d} \to \mathbb{R}^m$ as a linear fn. Apply w to X.

In what way is X_train the input data?

$$\text{X\_train} = \begin{array}{r} \\ \text{e.g. Bob, sample ①} \to \\ \text{e.g. Alice. sample ②} \to \\ \\ \text{e.g. Charlie. sample ⓜ} \to \end{array} \begin{array}{c} \text{feature ① e.g. height} \quad \text{feature ② e.g. weight} \quad \text{feature ⓓ e.g. size of their house} \\ \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} \end{pmatrix} \end{array}$$

output to predict, e.g. their income

$$\text{y\_train} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

m = no. of samples

d = no. of features.

$$w = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} \begin{array}{l} \leftarrow \text{weight attached to feature ①} \\ \\ \leftarrow \text{weight attached to feature ⓓ} \end{array}$$

When we multiply $X \cdot w$, we get

(sample index, feature index)

$w_1 \cdot x_{11} + w_2 \cdot x_{12} + \ldots + w_d \cdot x_{1d} = y_1$

$w_1 \cdot x_{21} + w_2 \cdot x_{22} + \ldots + w_d \cdot x_{2d} = y_2$

$\ldots$

Summary

$(\cdot)^T X$

- Operations on Vectors and Matrices
  - Dot-product, matrix inverse $X(\cdot)$
- Systems of Linear Equations $\mathbf{Xw} = \mathbf{y}$
  - Matrix-vector notation, linear dependency, invertible
  - Even-, over-, under-determined linear systems
- Set and Functions

Assignment 1 (week 6)
Tutorial 4

$m = d$

$(X^T X)^{-1} X^T$
$= X^{-1} (X^T)^{-1} X^T$

$\mathbf{X}$ is Square	Even-determined	m = d	One unique solution in general	$\hat{\mathbf{w}} = \mathbf{X}^{-1}\mathbf{y}$ $= X^{-1}$
$\mathbf{X}$ is Tall	Over-determined	m > d	No exact solution in general; An approximated solution	$\hat{\mathbf{w}} = \underbrace{(\mathbf{X}^T\mathbf{X})}_{d \times d}{}^{-1}\mathbf{X}^T\mathbf{y}$ Left-inverse
$\mathbf{X}$ is Wide	Under-determined	m < d	Infinite number of solutions in general; Unique constrained solution	$\hat{\mathbf{w}} = \mathbf{X}^T\underbrace{(\mathbf{X}\mathbf{X}^T)}_{m \times m}{}^{-1}\mathbf{y}$ Right-inverse

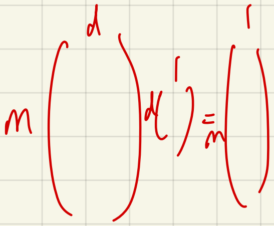
- Scalar and vector functions
- Inner product function
- Linear and affine functions

python package *numpy*
Inverse: *numpy.linalg.inv(X)*
Transpose: *X.T*

When solving for best w in $Xw = y$, m = no. of samples

d = no. of features

Case ①: $m > d$. Too many constraints.



→ cannot fit w to every constraint!

→ can find best approx. (minimises $\|Xw - y\|^2$) $Xw \approx y$.

→ how? use $\boxed{\text{Left Inverse}}$ set $\hat{w} = \underbrace{(X^TX)^{-1}}_{\text{provided... } X^TX \text{ is invertible}} X^T (Xw) = (X^TX)^{-1} X^T y_{\text{train}}$.

"reverses the transformation" of X

(exactly when X is injective.)

X: $m \times d$

X^T: $d \times m$

$\underbrace{X^TX}$: $d \times d$.

case ②: $m < d$. Too many features, too little data.

→ Too many possible solutions that give $Xw = y$.

→ think of X as a linear map. If X is surjective we use Right Inverse!

→ add more constraints (not mentioned)

→ use $\hat{w} = X^T \underbrace{(XX^T)^{-1}}_{\text{provided... } XX^T \text{ is invertible}} y$.

(exactly when X is surjective).

X: $m \times d$

X^T: $d \times m$

XX^T: $m \times m$.

---

dream case ③: $m = d$ AND X is invertible.

$$Xw = y$$

$$w = X^{-1}(Xw) = X^{-1}y$$

Linear Regression

"$y = wx + c$" , or "$X\_aug \cdot \hat{w} = y$"

$y = \underbrace{(1 \ x)}_{x_{aug}} \underbrace{\begin{pmatrix} c \\ w \end{pmatrix}}_{w_{aug}}$

e.g. $f(x_1, x_2) = w_0 \cdot 1 + w_1 x_1 + w_2 x_2$

where $f(\cdot)$ predicts y using x_1 & x_2.

Available data: X_train , y_train , X_test

what to do?

① Augment X : $X\_aug \longleftarrow \left(\begin{array}{c|c} 1 & \\ \vdots & X\_train \\ 1 & \end{array} \right)$

② determine shape of X_aug : m = no. of rows (samples)
d = no. of coln. (features)

③ Apply Left / Right Inverse to get linear weights for the linear model

$m > d$ (primal) : $\hat{w} \longleftarrow (X^T X)^{-1} X^T \cdot y\_train$ (Left inverse)

$m < d$ (dual) : $\hat{w} \longleftarrow X^T (X X^T)^{-1} \cdot y\_train$ (Right inverse)

use $X := X\_aug$

Note: $\hat{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix}$ — constant coeff (w_0); w_i is the weight for x_i for $X = (x_1, \dots, x_n)$

④ Apply linear model to X_test to predict y_pred

! need to augment X_test first!

ⓐ $X\_test\_aug \longleftarrow \left(\begin{array}{c|c} 1 & \\ \vdots & X\_test \\ 1 & \end{array} \right)$

ⓑ $y\_pred \longleftarrow X\_test\_aug \cdot \hat{w}_{aug}$.

Ridge Regression

What happens if $X^T X$ (resp. $X X^T$) is not invertible?

→ use $X^T X + \lambda I$ (resp. $X X^T + \lambda I$) instead with $\lambda > 0$ small. e.g. $\lambda = 0.0001$

→ this is called regularisation.

→ two purposes : ① makes $X^T X + \lambda I$ invertible
② biases the cost fn, to prefer $\hat{w}$ that is smaller

more precisely, s.t. $\|\hat{w}\|^2 = \hat{w}^T w$ is small

cost fn in primal form

original cost : $\min_w \|Xw - y\|^2 = \min_w (Xw - y)^T (Xw - y)$

regularised cost : $\min_w \|Xw - y\|^2 + \lambda \|w\|^2 = \min_w (Xw - y)^T (Xw - y) + \lambda w^T w$

How to get w that minimises the cost? : differentiate to get soln → Left Inverse. is now $\boxed{(X^T X + \lambda I)^{-1} X^T}$

cost fn in dual form (extra)

original cost : $\min_w \|Xw - y\|^2 = \min_w \sum_{i=1}^{m} (y_i - x^{(i)} w)^2$ where $X = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$

regularised cost : $\min_w \|Xw - y\|^2 + \lambda \|w\|^2 = \min_w \sum_{i=1}^{m} (y_i - x^{(i)} w)^2 + \lambda \sum_{j=1}^{d} w_j^2$, where $w = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}$

How to get (extra) w that minimises the cost? : need to apply Lagrangian duality. See KKT theory. This is why this is extra.

rough idea : get dual problem
solve dual problem.