TASK _ 27

$x^2 \neq x^{(2)}$ index

Machine learning

It's The science of getting computers to learn without being explicity programmed.

Applications of Machine learning AGI "artificial general intelligence"

Linear Regression Model Part 1

عبارة عن تمثيل خط بمجموعة من البيانات "Surpervise learning Model" لأنه يتنبأ بالبيانات

- يوجد منه نموذج اخر شائع "Classification Model" - يتنبأ بالفئات والفئات المتقطعة

→ gives Small number of possible outputs

بينما في حالة ال Regression Model → gives infinitely Many possible outputs

Notation

x: "input" Variable or "input" Feature

y: "output" " " "output" target Variable

m: number of training examples

(x, y): Single training example

$(x^{(i)}, y^{(i)})$: i^{th} training example "1st, 2nd, 3rd ...)

feature $x \rightarrow \boxed{f} \rightarrow \hat{y}$ → Prediction "estimated y"

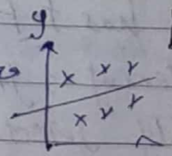
Model → $f(x)$

features, targets → training set → learning algorithm → $\boxed{f}$ → function

How to represent f?

→ $f_{w,b}(x) = wx + b$ معادلة خط مستقيم

$f_{w,b}(x^{(i)}) = \hat{y}^{(i)}$



$F(x) = wx + b$

Parameters

Slope w, b تقاطع y

Cost function

"squared error cost function"

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(\underbrace{f_{w,b}(x^{(i)}) - y^{(i)}}_{\hat{y}^{(i)}}\right)^2$$

error

intuition

Machine learning

→ Field of study that gives computers the ability to learn without being explicitly programmed

Machine learning algorithms

- Supervised learning → Most in real-world applications → rapid advancements
- unsupervised "
- Recommender systems
- Reinforcement learning

input (x)	output (y)	Applicat
email	Spam (0/1)	Spam filter
English	arabic	Machine trans
ad, user info	click (0/1)	Online adverts

1) Supervised Learning $\underset{\text{input}}{x} \rightarrow \underset{\text{output}}{y}$

* classification → Predict categories → Small number of possible outputs
  ↳ two or more inputs

Surpervised learning
learns from being given "right answers"

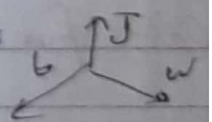
Regression
↳ Predict a number
infinitely many possible outputs

Classification
Predict categories
Small number of possible outputs

2) unsupervised learning → find something interesting in unlabeled data

Clustering: example DNA microarray

Data only comes with inputs x, but not output labels y.
Algorithm has to find structure in the data
↳ clustering: group similar data points together
↳ Anomaly detection: find unusual data points
↳ Dimensionality reduction: Compress data using fewer numbers



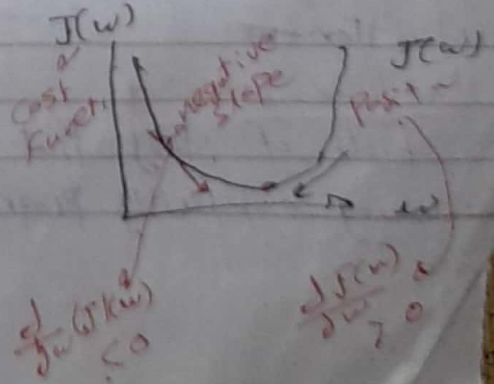
Have some function $J(w,b)$ → want $\min J(w,b)$ → الهدف هو الوصول للـ
لأدنى الـ cost funct gradient descent

Gradient descent algorithm

$$w = w - \alpha \boxed{\frac{d}{dw} J(w,b)}$$

→ α: learning rate
↳ Derivative

$$b = b - \alpha \frac{d}{db} J(w,b)$$



$w = w - \alpha \frac{d}{dw} J(w)$ → If α is too small

Gradient descent may be slow

→ If α is too large → " " " → overshoot

→ never reach minimum

Can reach local minimum with fixed learning rate α

→ fail to converge

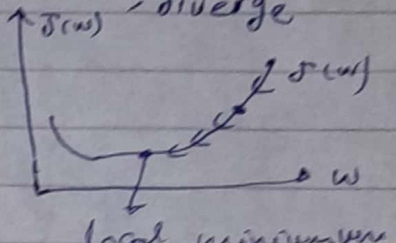
- diverge

Near a local minimum

↳ Derivative becomes smaller

↳ update steps become smaller

→ Can reach minimum without decreasing learning rate



$$\boxed{f_{w,b}(x^{(i)}) = wx^{(i)} + b}$$

Gradient descent for linear regression

linear regression model

$f_{w,b}(x) = wx + b$

Cost function

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} (f_{w,b}(x^{(i)}) - y^{(i)})^2$$

Gradient descent algorithm

repeat until convergence { $w = w - \alpha \frac{\partial}{\partial w} J(w,b) \to \frac{1}{m} \sum_{i=1}^{m} (f_{w,b}(x^{(i)}) - y^{(i)})$

$b = b - \alpha \frac{\partial}{\partial b} J(w,b) \to \frac{1}{m} \sum_{i=1}^{m} (f_{w,b}(x)^{(i)} - y^{(i)})$

"Batch": Each step of gradient descent uses all the training examp