stanCode

SC201 Lecture 3

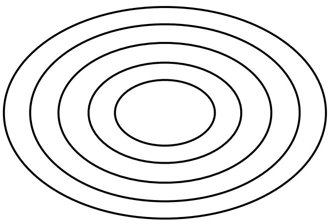
Gradient Descent h(θ, b) = ________

< ________ Gradient Descent > (BGD)

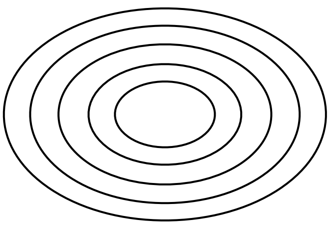
θ = θ - α (________________________)

b = b - α (________________________)

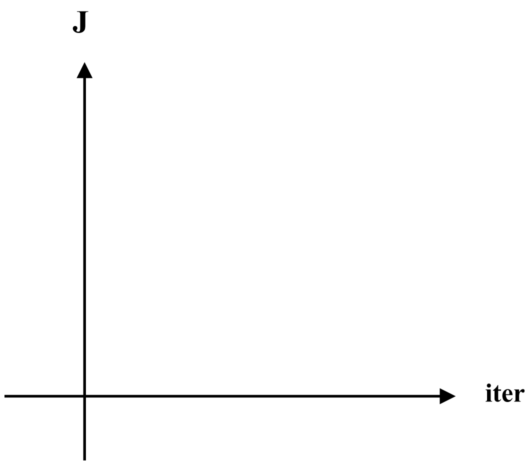
- 往 ______________________________ 的方向走去



Steepest Descent



Non-Steepest Descent

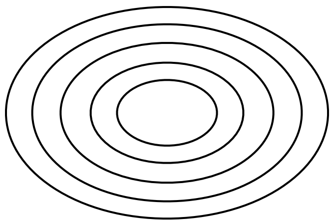


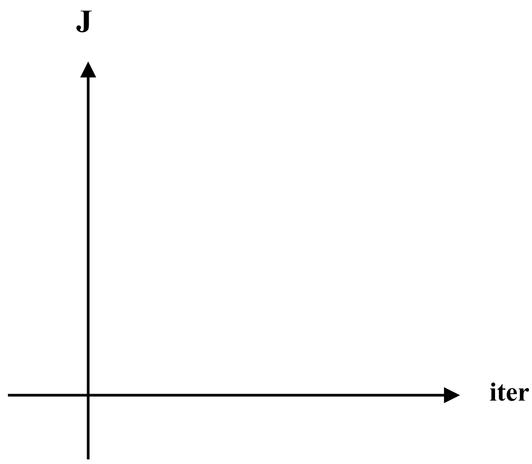
< ____________ Gradient Descent > (SGD)

θ = θ - α (________________________)

b = b - α (________________________)

- 往 ______________________________ 的方向走去





Hypothesis Function (Model)

<degree1 polynomial>

$$J = \frac{1}{2m}\sum_{i=1}^{m}(\theta Xi + b - yi)^2$$

$\Rightarrow$ $\boxed{J = \qquad\qquad}$

$\Rightarrow$ $\boxed{\frac{dJ}{d\theta} = \qquad\qquad}$

<degree2 polynomial>

$$J = \frac{1}{2m}\sum_{i=1}^{m}(\theta' xi^2 + \theta xi + b - yi)^2$$

From Learning to Predicting

<degree1 polynomial>

θ = θ - α(dJ_dθ)

b = b - α(dJ_db)

Learn Parameters

prediction = ______________________

<degree2 polynomial>

θ' = θ' - α(dJ_dθ')

θ = θ - α(dJ_dθ)

b = b - α(dJ_db)

Learn Parameters

prediction = ______________________

Supervised Learning

Input data has ______________________

< Regression Problem > ⟶ **< Classification Problem >**

< Regression Problem >

Output is ______________

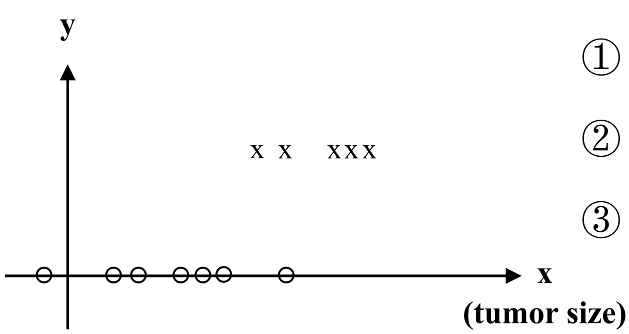
- Housing Price
- Steering wheel angle

< Classification Problem >

Output is ______________

- Benign/ Malignant Tumor
- Good/ Bad Review

Problems of using linear regression on binary classification



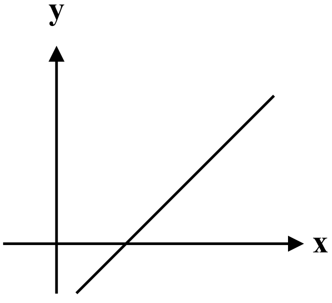
① ______________________ not defined.

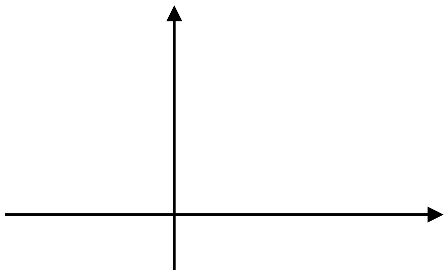
② ______________ doesn't follow linear relationship

③ sensitive to ______________

linear regression → logistic regression

$< k = \theta x + b >$ ⟶ $< h = \dfrac{1}{1 + e^{-k}} >$





① Bounded between ___ and ___

② Changes faster around ______________

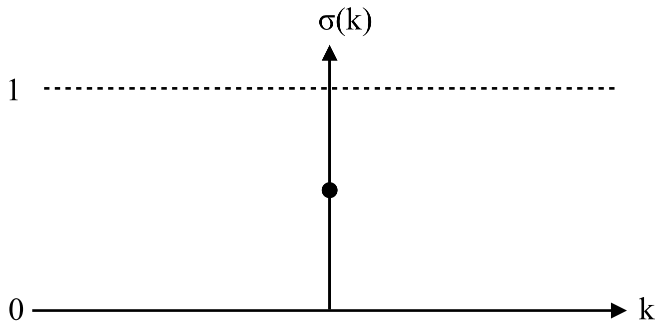
③ ___________ are easy to predict!

④ Can be ____________________________

< Sigmoid function >

$$\text{sigmoid}(k) = \sigma(k) = \frac{1}{1 + e^{-k}}$$

where k = _____________



- when k is ∞ (huge tumor)

  $\Rightarrow \dfrac{1}{1 + e^{-\infty}} =$ _____

- when k is -∞ (tiny tumor)

  $\Rightarrow \dfrac{1}{1 + e^{\infty}} =$ _____

- when k is 0 (the threshold for good/bad tumor)

  $\Rightarrow \dfrac{1}{1 + e^{0}} =$ _____

if k ___ 0 → σ(k) ___ 0.5 → predict ①

if k ____ 0 → σ(k) ____ 0.5 → predict ⓪

Supervised Learning

< Regression Problem >

_______________ regression **h = θx**

< Classification Problem >

_______________ regression $\mathbf{h = \sigma(\theta x) = \dfrac{1}{1 + e^{-(\theta x)}}}$

Cost Function

< Regression Problem >

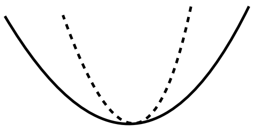
J = _____________________

< Classification Problem >

J = _____________________

Convex Function

$$\frac{1}{2m}\sum_{i=1}^{m}(\theta Xi - yi)^2$$



bowl shaped

There is ________________ minimum

(local min ___ global min)

Non-convex Function

$$\frac{1}{2m}\sum_{i=1}^{m}(\sigma(\theta Xi) - yi)^2$$

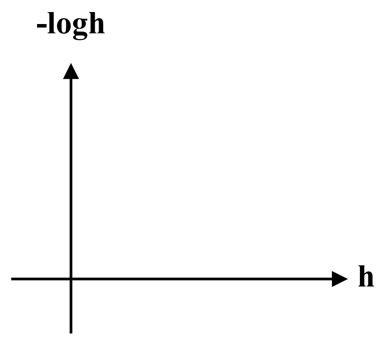
Depends on ______________

________________ minimums

(local min ___ global min)

Loss Function for Logistic Regression

L = ________________________________ (where hi = σ(θxi))



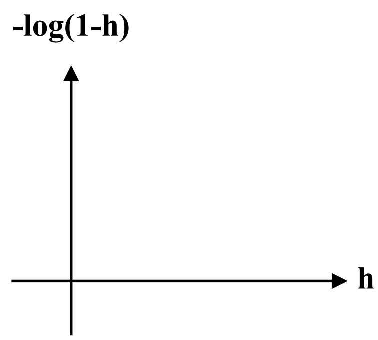
< if yi == 1 >

L = ___________

如果hi預測 0.85 ⇒ L = -log(____) = ______

如果hi預測 0.2 ⇒ L = -log(____) = ______

When yi == ____ ⇒ loss function is convex



< if yi == 0 >

L = ___________

如果hi預測 0.85 ⇒ L = -log(____) = ______

如果hi預測 0.2 ⇒ L = -log(____) = ______

When yi == ____ ⇒ loss function is convex

Cost Function for Logistic Regression

J = ________________________________

Gradient Descent for Logistic Regression

θ = ________________________________

Gradient Descent for Linear Regression

θ = ________________________________

$$\boxed{\frac{dJ}{d\theta}}$$

J = ____

hi = _______________

ki = _______________

$$\frac{dJ}{d\theta} = \boxed{\frac{\quad}{\quad}}\ \boxed{\frac{\quad}{\quad}}\ \boxed{\frac{\quad}{\quad}}$$

$$\frac{dJ}{dhi} =$$

$$\frac{dhi}{dki} =$$

$$\frac{dki}{d\theta} =$$

$$\frac{dJ}{d\theta} =$$