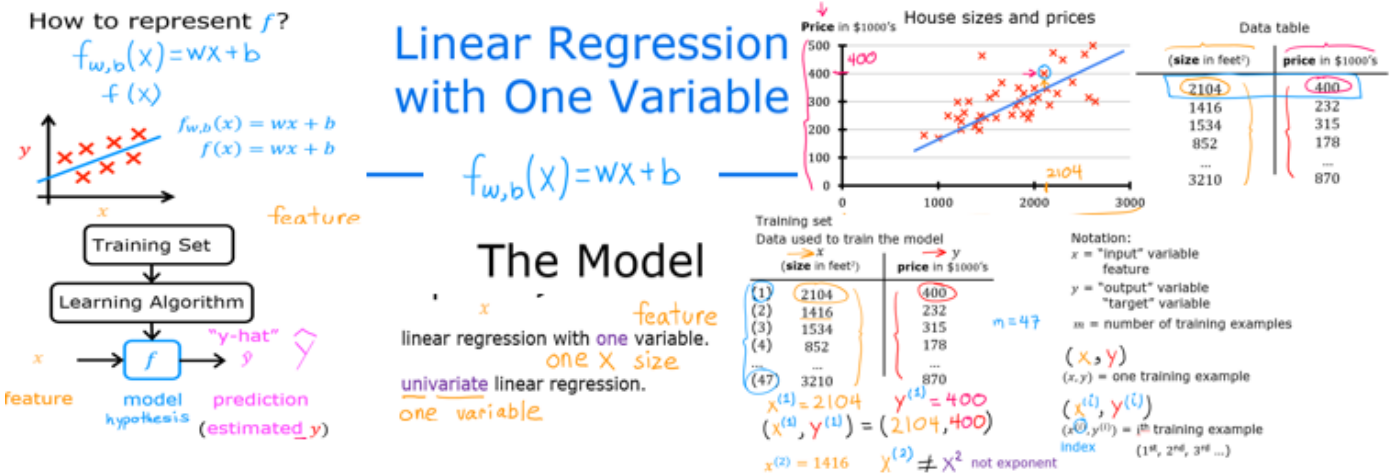
Regressione lineare

Tramite la *regressione lineare* univariata, viene determinato matematicamente l'output della funzione dato l'input, o se vogliamo dirla in altro modo, viene calcolata la Y in funzione della X. (tipicamente determinato la funzione che meglio fitta di valori X e Y)

La RL serve per determinare i valori della funzione che meglio soddisfano l'andamento di un fenomeno relativo ai valori appartenenti all'insieme delle fetures/labels (x,y)

La funzione si può rappresentare come y = b + wx dove **b** è “intercetta” (ovvero il delta y rispetto allo zero detto anche “bias” in inglese) mentre la **w** è il “coefficiente angolare. (ovvero l'inclibazione della retta detto anche ”weight" in inglese o “slope” ovvero pendenza nel caso della funzione)

Nel Machine Learning (ML) i parametri “w” e “b” sono detti anche coefficienti o pesi.

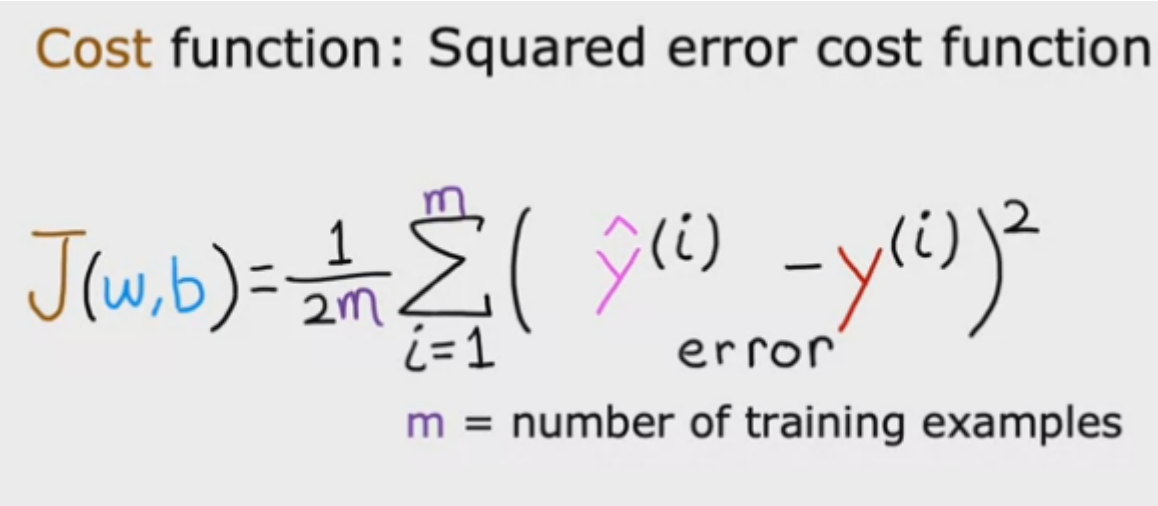


Costo della funzione

La Cost Function (CF) nella pratica confronta il valore “predetto” y^ e il valore di training y, nella pratica è una differenza tra i due valori che definiamo come “errore”, ponendo il delta al quadrato, ovvero:

Sommatoria di (y^ - y)^2 per tutti i valori “m” del training set; diviso per “m” ovvero, viene ritornato la media degli errori -> questo si chiama “metodo dei minimi quadrati”

L'intento è quindi quello di trovare il valore minimo della funzione di costo.



Si tratta quindi di trovare la funzione che minimizza l'errore.

Nell'immagine sotto riportata a sx vengono visualizzate le funzioni ottenibili al variare del permarametro “w”

NB: per semplicità l'esempio pone b= 0 per rendere il grafico più facilemente interpretabile perchè in questo modo il gradico è 2D, se considerassimo anche

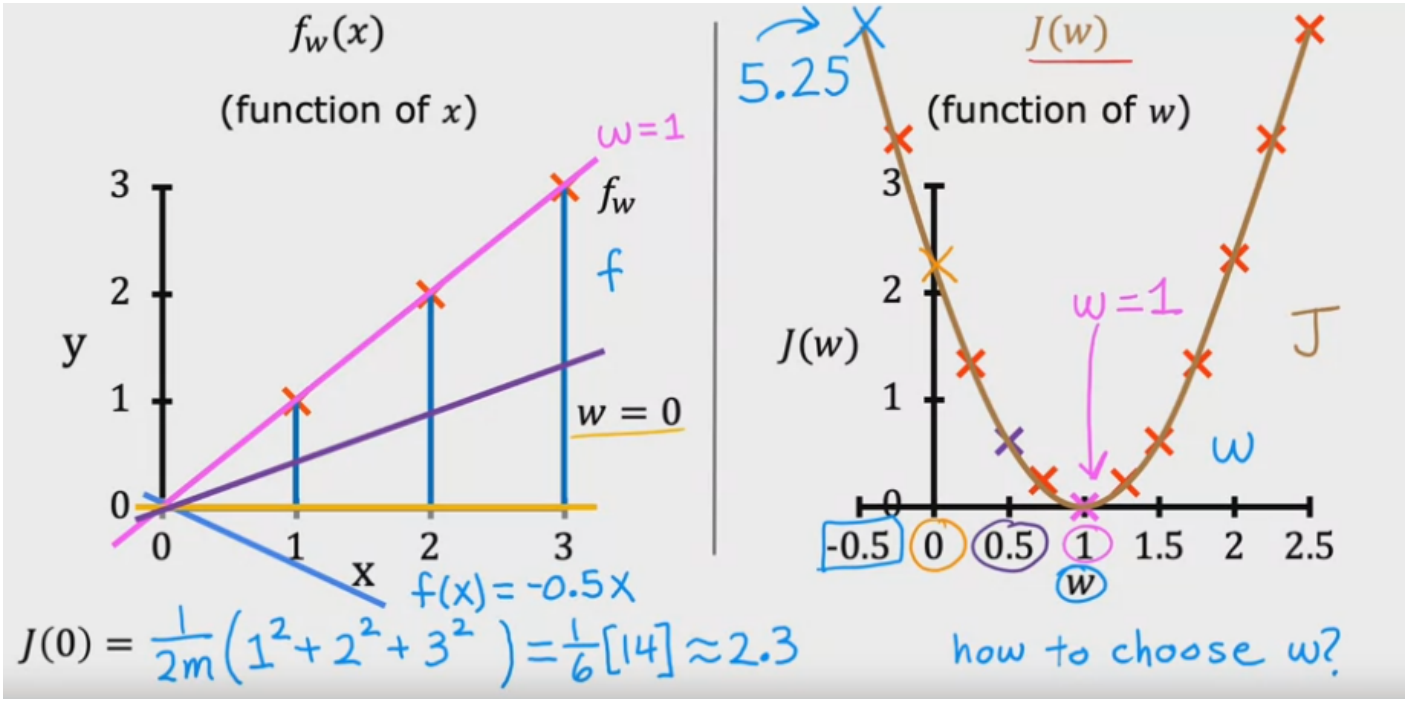
il valore b, allora il grafico sarebbe 3D e quindi un pò più difficile da leggere. (vedi es. grafico 3D riportato in pagina)

Si può notare che nella parte sx al varia della funzione si ottengono diversi valori J (ovvero errore) che, se rappresenti graficamente disegnano una curva

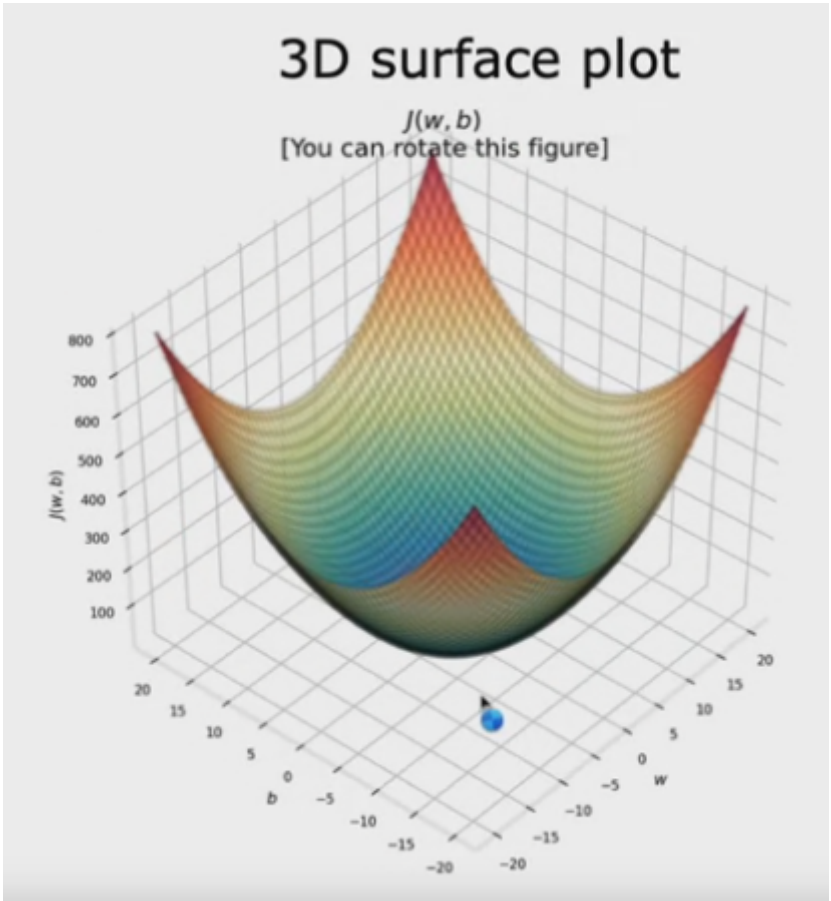
dove l'errore minimo si trova quando, per es. nel caso specifico, w vale 1. (grafico a DX)

Il grafico a DX disegna sulle ordinate (y) l'errore mentre sulle ascisse (x) il valore w ottenuti dall'applicazione del metodo

dei minimi quadrati. (vedi grafico a SX)



Nel caso in cui considerassimo anche il parametro “b", il grafico del cost function sarebbe renderezzito in 3D, come sotto rappresentato.



Considerazioni finali

Per determinare il “costo della funzione minimo” o l'errore minimo bisogna determinare i parametri “w” e “b” che avvicinano la funzione al set di dati.

Per determinare questi parametri in maniera algoritmica viene utilizzata la tecnica matematica denominata “gradiant descent” o “discesa del gradiente” analizzata nel paragrafo successivo.

Discesa del gradiente

La discesa del gradiente (GD) è un modo sistematico per determinare i parametri “w” e “b” in modo che minimizzino

il "costo della funzione" (l'errore della funzione)

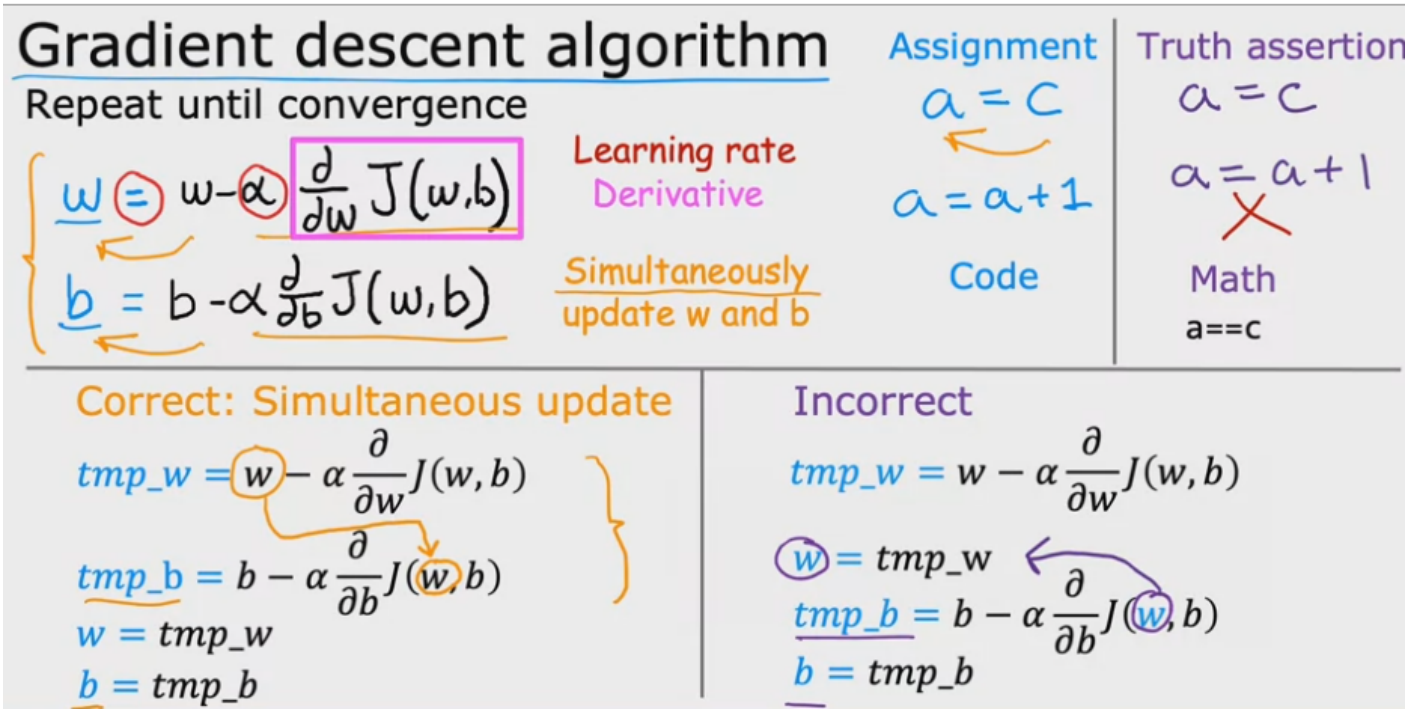
Ricordo che stiamo parlando della funzione i cui coefficienti (w e b) contribuiscono a determinare

i valori che più si avvicinano al set di dati passati in input (features e labels)

Quindi J(w,b) -> costo della funzione (ovvero la sommatoria dei delta dati da labels - labels calcolate con i coifficienti w)

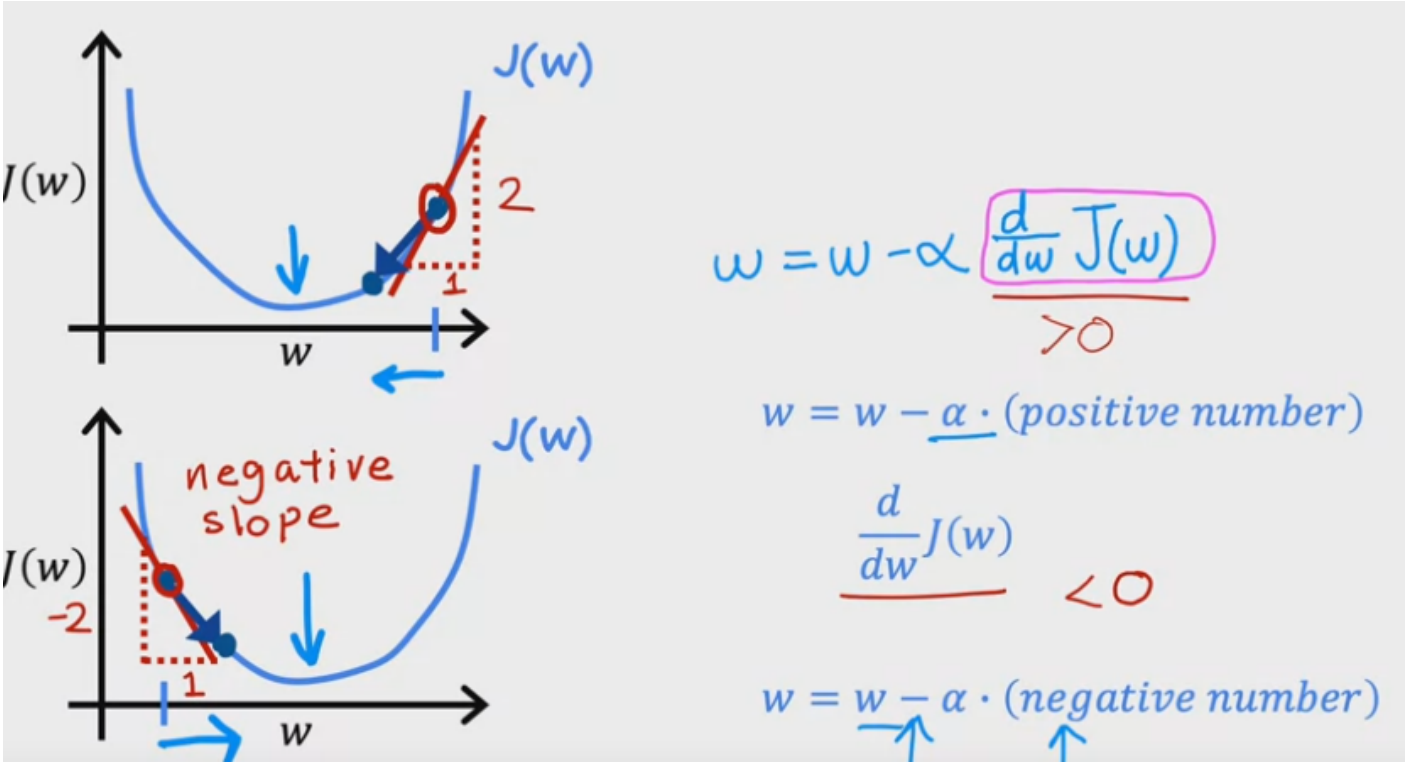
Tramite questo metodo si procede per “piccoli passi” al fine di trovarel l'errore minimo J(x,b), nota che

possono esistere più minimi (detti anche “minimi locali”) che dipendono dal valore di partenza che in genere è randomico.



Nella pratica si tratta di looppare i valori “w” e "b" calcolati ad ogni iterazione del ciclo. (vedi immagine sopra)
Il valore di w è quindi settato a = “w” meno un parametro “alpha” (detto anche learning rate, valore piccolo compreso tra 0 e 1, es. 0,001) moltiplicato
per la derivata del costo della funzione J(w,b).
Stessa cosa per il parametro b.

Il valore di “learing rate” (LR) è indica la grandezza del “passo”, ovvero la velocità con la quale saliamo i discendiamo il grafico
del costo della funzione.
Attenzione che andare veloce può causare un “salto” eccessivo e farci perdere il punto di minimo.



Sopra viene indicato il calcolo della derivata parziale rispetto al valore “w”. Vengono rappresentati due casi, il primo

dove viene preso a caso un valore “w” a DX dal minimo; In questo caso la derivata parziele ritornerà un numero positivo

in quanto, la derivata indica l'inclinazione della tangente passante per il punto scelto sulla curva avente per coordinate (w1,j(w1))

e in questo caso indica che la tangente è ascendente.

La derivata del costo della funzone indica se la funzione è a un minimo (anche locale) oppure se si trova in discesa o in salita. In pratica

restituisce la pendenza (o la direzione) del punto tangente alla funzione di costo nelle coordinate indicate.

Ora il nuovo “w” viene calcolato sotrraendo il " LR x lla derivata" al valore “w” , essendo positivo, diminuirà il valore finare di “w”.

Stessa cosa, ma invertita di segno in quanto il valore di w preso a SX del minimo è discendete. Quindi, il valore della derivato

del costo della funzione viene negato in quanto la formula sottrae sempre il valore della derivta parziale del costo di “w” e di “b”.

Idem per la variabile “b” fino a quando i valori convergono intorno al minimo della funzione.

Learning rate

Il valore di “learning rate” (LR) nella pratica serve per fare “lo step” nella direzione del minimo in cui valore è dato dal vosto della funzione.

LR però non può essere un valore troppo piccolo in quanto rallenterebbe in maniera importante la determinazione del minimo e, non può

essere nemmeno troppo grande in quanto rischia di far saltare il punto di minimo, sia esso locale che globale.

Per questo motivo il modo migliore per settare il parametro alpha (detto anche LR) è gestire dinamicamente il valore da algoritmo in modo

che sia un valore relativamente grande se la pendenza (slope) dato dalla derivata nel punto della funzione di costo è elevato, piccolo se lo

slope tende allo zero in quanto significa che siamo prossimi al minimo.

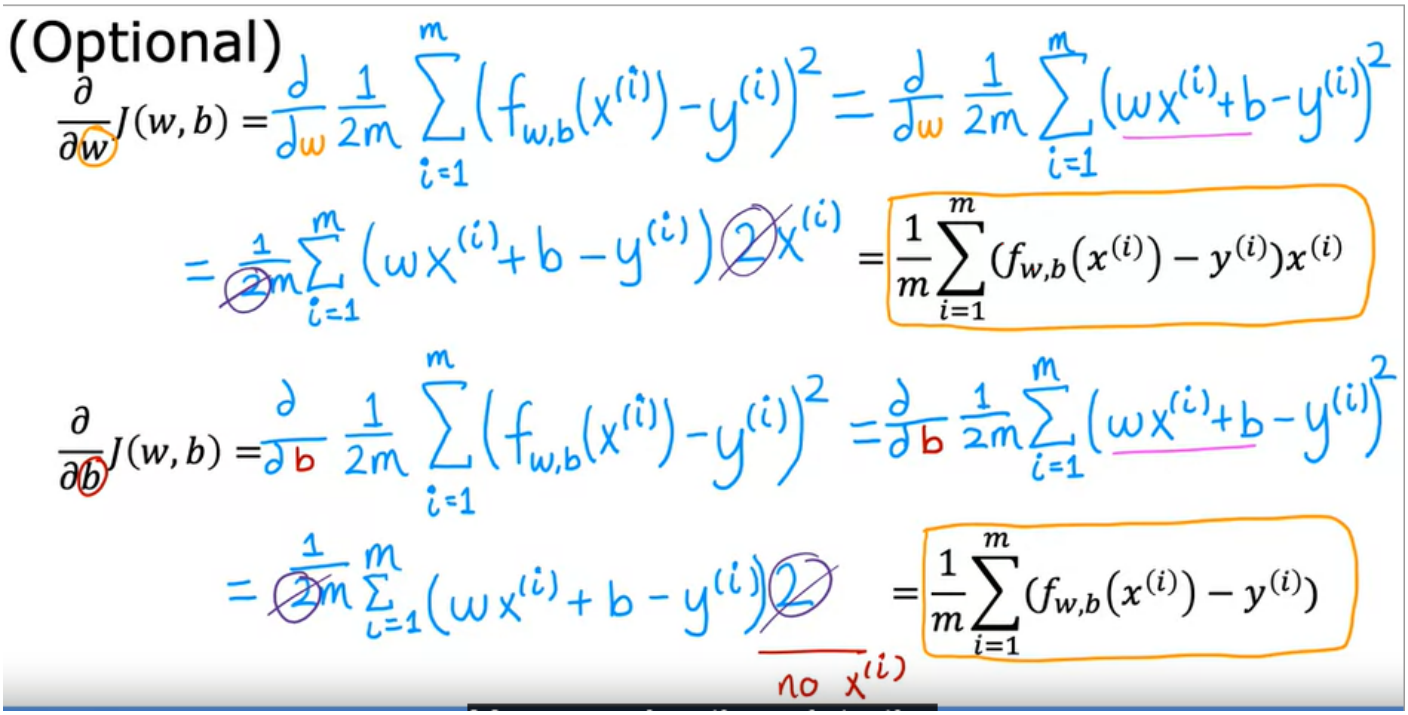
Di seguito viene mostrato il calcolo delle deritavate parziali rispetto a “w” e rispetto a “b" per determinare il minumo del costo della funzione.

NOTA di calcolo, si tratta di applicare il teorema della derivata della funzione composta *(detta anche "chain rule)* rispetto a **w** e rispetto a **b** ovvero:

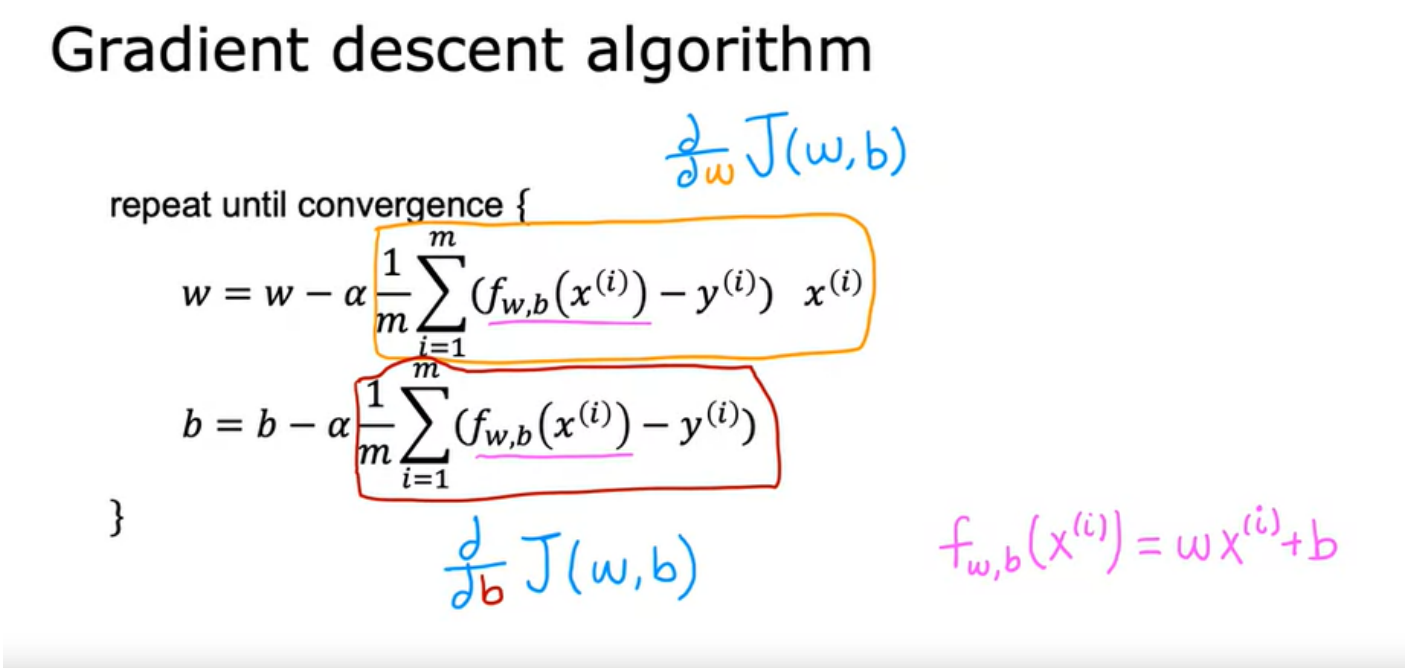
$$\frac{d}{dx}h(x) = g'(f(x)) \cdot f'(x)$$

In parole povere, la derivata della funzione composta h(x)=g(f(x)) è data dalla derivata della funzione più esterna, *con argomento invariato*, moltiplicata per la derivata della funzione più interna. Con funzione più esterna si intende l'ultima funzione che si applica nella composizione (per noi la g) mentre la più interna è la prima che si applica (f).

che applicando il tutto alla funzione di costo si ottiene:



Quindi l'algoritmo di calcolo della discesa del gradiente si calcola come:



Conclusione

In conclusione, la regressione lineare consente di determinare il valore minimo relativo al costo della funzione.

Il costo della funzione ritorna la sommatoria degli scostamenti tra la funzione (ad una o più variabili) e l'insieme dei valori (campioni)

labels-features al variare dei coefficiente “w” e “b”.

Al variare di questi coefficienti si genera un insieme di valori dove ciascun valore è un totale che se rappresentato

graficamente, disegna un grafico in 2D (se varia solo un valore) in 3D se variano sia “w” che “b”, del quale dobbiamo trovare

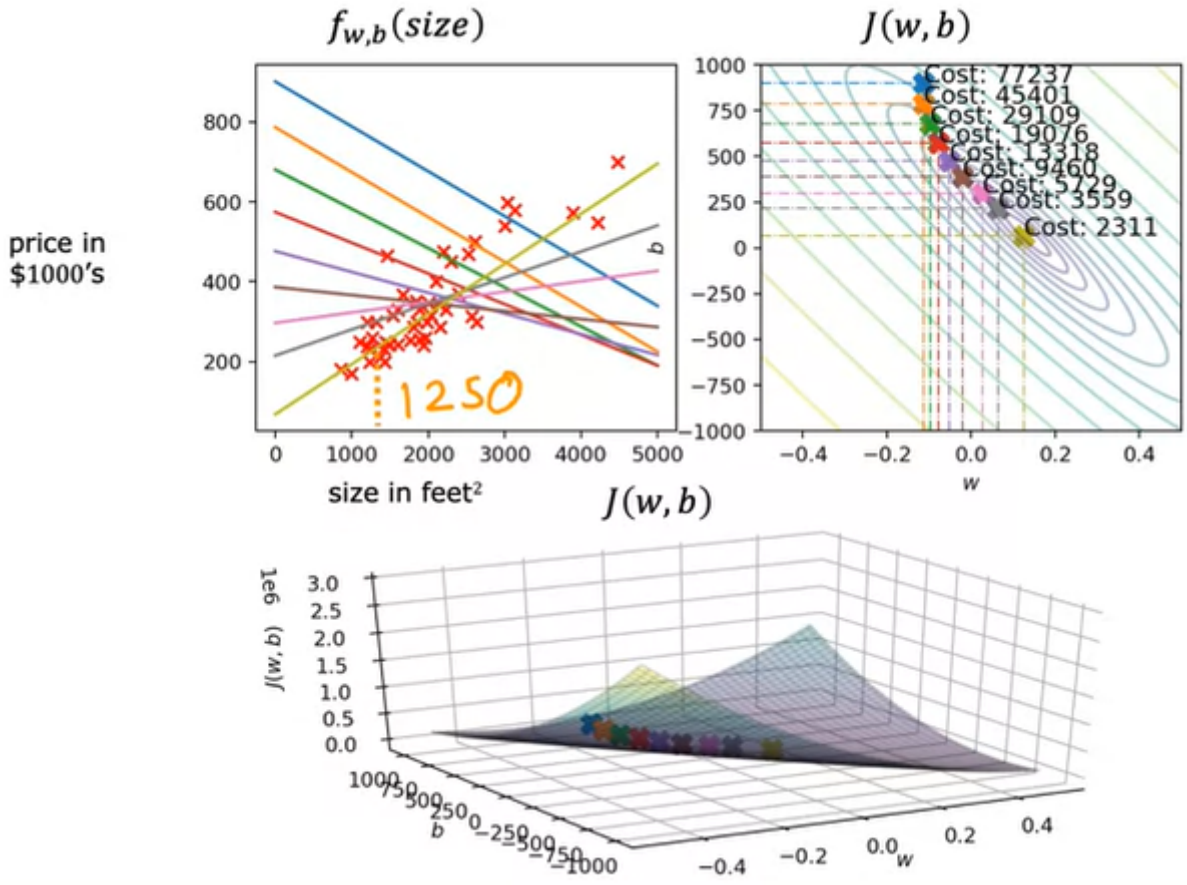
il minimo per identificare la funzione che minimizza l'errore.

Per trovare l'errore minimo si utilizza il metodo della discesa del gradiende, che consente tramite l'utilizzo delle derivate parziali

nel punto iesimo, di avvicinarsi progressivamente al “minimo locale” oppure al “minimo assoluto” avanzando tramite un passo definito “**learing rate**”.

Di seguito un esempio di come al variare dell'intercetta e del coefficiente angolare, andando con passo (LR) la funzione

giunge al suo costo minimo.



Revision #1
Created 2023-03-11 16:44:16 UTC by marco
Updated 2024-10-13 15:09:54 UTC by marco