# Report Dettagliato della Lezione con Esempi Pratici

Questo documento riassume in modo dettagliato, completo e con esempi pratici i concetti trattati durante la lezione di oggi. Ogni sezione è sviluppata per fornire un'ampia comprensione dei concetti, con spiegazioni intuitive, tecniche e applicazioni pratiche.

## 1. Funzione Lineare Compressa

La funzione lineare compressa è una combinazione di due passaggi: calcolo di un punteggio grezzo tramite una funzione lineare e compressione del punteggio in un intervallo interpretabile, solitamente tra 0 e 1, usando funzioni di attivazione come sigmoid o softmax.

### Esempio Pratico

Supponiamo di voler prevedere se un'email è spam o non spam in base a:  
- Numero di parole (x1 = 50).  
- Numero di link (x2 = 3).  
- Presenza della parola 'gratis' (x3 = 1).  
Con i pesi w1 = 0.1, w2 = 1.2, w3 = 2.0, e bias b = -4, il punteggio grezzo è:  
z = (0.1 \* 50) + (1.2 \* 3) + (2.0 \* 1) - 4 = 6.6  
Applichiamo la funzione sigmoid per calcolare la probabilità:  
P(Spam) = 1 / (1 + e^(-6.6)) ≈ 0.9986  
La probabilità è del 99.86%, quindi l'email è classificata come spam.

## 2. Regressione Logistica

La regressione logistica è un algoritmo di classificazione binaria che utilizza una funzione lineare e la funzione sigmoid per trasformare un punteggio grezzo in una probabilità. Se la probabilità supera una soglia (tipicamente 0.5), assegna l'input alla classe positiva.

### Esempio Pratico

Consideriamo un modello per prevedere se un cliente acquisterà un prodotto. Le caratteristiche sono:  
- Età (x1 = 30).  
- Reddito annuo in migliaia (x2 = 50).  
Con pesi w1 = 0.8, w2 = 1.5, bias b = -60, il punteggio grezzo è:  
z = (0.8 \* 30) + (1.5 \* 50) - 60 = 75.  
La probabilità è calcolata con la sigmoid:  
P(Acquisto) = 1 / (1 + e^(-75)) ≈ 1.  
Quindi il cliente è classificato come molto probabile acquirente.

## 3. Discesa del Gradiente

La discesa del gradiente è un algoritmo utilizzato per ottimizzare una funzione obiettivo (ad esempio, l'errore del modello). Calcola il gradiente della funzione rispetto ai parametri del modello e aggiorna i parametri muovendosi nella direzione opposta al gradiente.

### Esempio Pratico

Supponiamo di avere una funzione obiettivo J(w) = (w - 2)^2. Il gradiente è ∇J(w) = 2(w - 2). Iniziamo con w = 0 e un tasso di apprendimento η = 0.1.  
- Iterazione 1: w = w - η \* ∇J(w) = 0 - 0.1 \* 2(0 - 2) = 0.4.  
- Iterazione 2: w = 0.4 - 0.1 \* 2(0.4 - 2) = 0.72.  
Ripetendo questo processo, w converge a 2, che è il minimo della funzione.

## 4. Discesa del Gradiente Stocastica (SGD)

La discesa del gradiente stocastica aggiorna i parametri del modello utilizzando un solo esempio alla volta, rendendo l'algoritmo più veloce ma meno stabile rispetto alla discesa del gradiente standard.

### Esempio Pratico

Consideriamo un dataset con due esempi: x1 = 1, y1 = 2 e x2 = 2, y2 = 4. Supponiamo che la funzione obiettivo sia J(w) = (y - wx)^2. Iniziamo con w = 0 e η = 0.1.  
- Per x1: w = 0 - 0.1 \* ∇J(0; 1, 2) = 0.4.  
- Per x2: w = 0.4 - 0.1 \* ∇J(0.4; 2, 4) = 0.64.  
L'algoritmo aggiorna i parametri ad ogni esempio, rendendolo più veloce per dataset grandi.

## 5. Differenze tra Batch Gradient Descent e SGD

Batch Gradient Descent calcola il gradiente su tutto il dataset, risultando preciso ma lento. SGD utilizza un esempio alla volta, rendendolo più veloce ma meno stabile. Una via di mezzo è il Mini-Batch Gradient Descent, che utilizza piccoli gruppi di esempi per ogni aggiornamento.

## Conclusione

Gli argomenti trattati oggi coprono tecniche fondamentali per la classificazione e l'ottimizzazione nei modelli di machine learning. La comprensione di queste tecniche è essenziale per sviluppare modelli efficienti e accurati.