Analisi di Immagini e Video (Computer Vision)

Giuseppe Manco

Outline

- Preliminari
- Classificazione
  - Esempio: regressione logistica
- Quali features?

Crediti

- Slides adattate da vari corsi e libri
  - Analisi di Immagini (F. Angiulli) – Unical
  - Intro to Computer Vision (J. Tompkin) – CS Brown Edu
  - Computer Vision (I. Gkioulekas) - CS CMU Edu
  - Cmputational Visual Recognition (V. Ordonez), CS Virgina Edu
  - Pattern Recognition and Machine Learning (C. Bishop, 2005)
  - Deep Learning (Bengio, Courville, Goodfellow, 2017)

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- f di forma non nota
- x campionato da un dominio X e y campionato da un dominio Y
- Tipicamente, $x \in \mathbb{R}^n$

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- f di forma non nota
- x campionato da un dominio X e y campionato da un dominio Y
- Tipicamente, $x \in \mathbb{R}^n$
- Supervised Learning
  - Apprendi f da $D \subset X \times Y$

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- f di forma non nota
- x campionato da un dominio X e y campionato da un dominio Y
- Tipicamente, $x \in \mathbb{R}^n$
- Supervised Learning
  - Apprendi f da $D \subset X \times Y$
- Unsupervised Learning
  - Apprendi f da $D \subset X$

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- f di forma non nota
- x campionato da un dominio X e y campionato da un dominio Y
- Tipicamente, $x \in \mathbb{R}^n$
- Supervised Learning
  - Apprendi f da $D \subset X \times Y$
- Semi-supervised Learning
  - Apprendi f da $D = (D_1, D_2)$ dove $D_1 \subset X \times Y$ e $D_2 \subset X$

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- f di forma non nota
- x campionato da un dominio X e y campionato da un dominio Y
- Tipicamente, $x \in \mathbb{R}^n$

$$f(x) \equiv f(x;\theta) \equiv f_\theta(x)$$

- $\theta \in \Theta$ parametro scelto dal parameter space Θ (il linguaggio)

Inferenza vs. Learning

- Stima/Apprendimento: Selezionare
  - I parametri più appropriati
  - Una distribuzione sui parametri
  - Un insieme di distribuzioni
- Inferenza:
  - Predizioni
  - Statistiche
  - Valori attesi
  - Margini di un modello statistico

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- f di forma non nota
- x campionato da un dominio X e y campionato da un dominio Y
- Tipicamente, $x \in \mathbb{R}^n$

$$f(x) \equiv f(x;\theta) \equiv f_\theta(x)$$

- $\theta \in \Theta$ parametro scelto dal parameter space Θ (il linguaggio)
- Cos'è y?

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- Cos'è y?
- Una classe

cat $= f($  $)$

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

- Cos'è y?
- Un insieme finito di regioni

 $= f($  $)$

Supervised, unsupervised e semi-supervised learning

$$y = f(x)$$

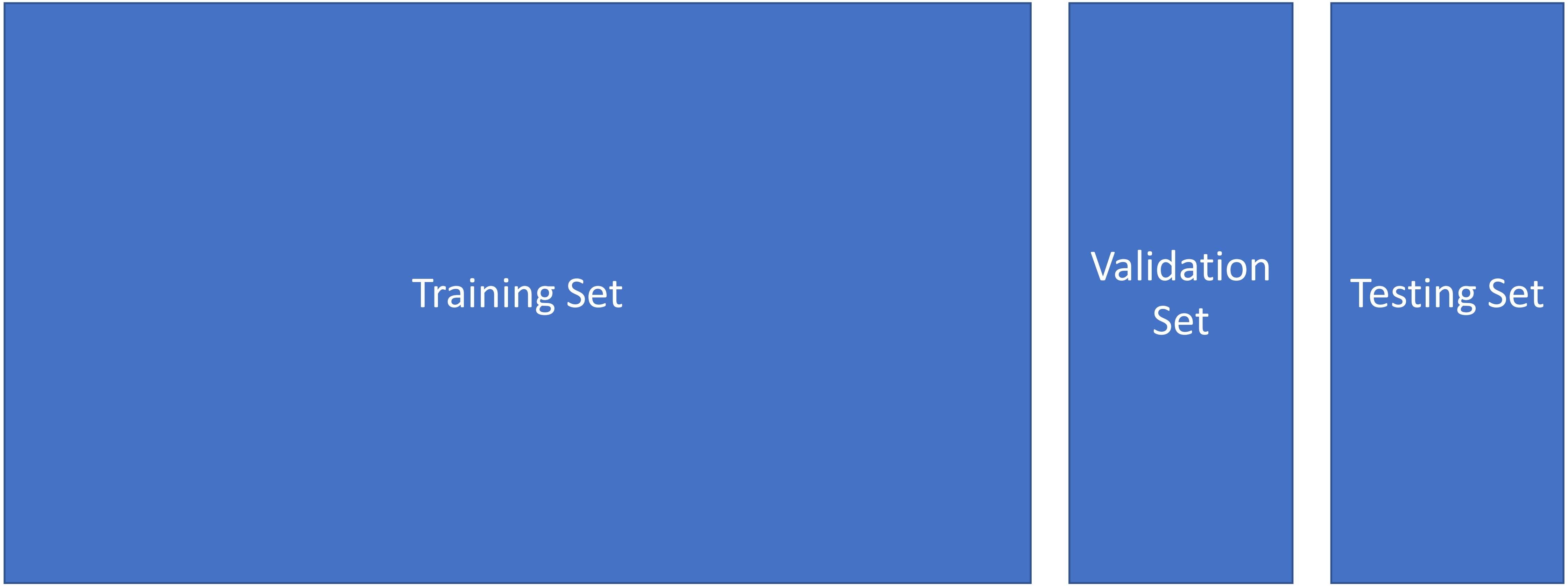
- Cos'è y?
- Segmenti

 $= f($  $)$

Inferenza vs. Learning

- Stima/Apprendimento: Selezionare
  - I parametri più approriati
  - Una distribuzione sui parametri
  - Un insieme di distribuzioni
- Inferenza:
  - Predizioni
  - Statistiche
  - Valori attesi
  - Margini di un modello statistico

Training, Validation, Test Sets



D

Training, Validation (Dev), Test Sets

D_{train}

Usato nella fase di learning

Training, Validation (Dev), Test Sets

Usato nella fase di valutazione

D_{test}

Classificazione

Classificazione

Training Set

cat

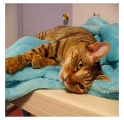
dog

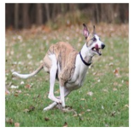
cat

.
.
.

bear

Test Set





.
.
.

Classificazione

Training Set

$x_1 = [$  $]$ $y_1 = [\text{cat}]$

$x_2 = [$  $]$ $y_2 = [\text{dog}]$

$x_3 = [$  $]$ $y_3 = [\text{cat}]$

⋮

$x_n = [$  $]$ $y_n = [\text{bear}]$

Classificazione

Training Set

inputs	targets / ground truth	Predizioni
$x_1 = [x_{11} \quad x_{12} \quad x_{13} \quad x_{14}]$	$y_1 = 1$	$\hat{y}_1 = 1$
$x_2 = [x_{21} \quad x_{22} \quad x_{23} \quad x_{24}]$	$y_2 = 2$	$\hat{y}_2 = 2$
$x_3 = [x_{31} \quad x_{32} \quad x_{33} \quad x_{34}]$	$y_3 = 1$	$\hat{y}_3 = 2$
. . .		
$x_n = [x_{n1} \quad x_{n2} \quad x_{n3} \quad x_{n4}]$	$y_n = 3$	$\hat{y}_n = 1$

Dato il parametro θ possiamo calcolare $\hat{y}$

$$\hat{y} = f_\theta(x)$$

Problema:
Come trovare il parametro θ?
Soluzione:
Funzione di costo

$$loss = \sum_i Cost(y_i, \hat{y}_i)$$

Modello lineare di classificazione

- Caso semplice: classificazione binaria
  - $y \in \{0,1\}$
    - (o $y \in \{1,2\}$, $y \in \{-1,1\}$, …)

- Il decision boundary tra due classi è un iperpiano nello spazio delle features
- Le due regioni sono separate dall’iperpiano

$$w_1 x_1 + w_2 x_2 + \cdots w_m x_m = b$$

Regressione logistica

Training Data

inputs

targets /
labels /
ground truth

$x_1 = [x_{11} \quad x_{12} \quad x_{13} \quad x_{14}]$ $\quad y_1 = 1$

$x_2 = [x_{21} \quad x_{22} \quad x_{23} \quad x_{24}]$ $\quad y_2 = 2$

$x_3 = [x_{31} \quad x_{32} \quad x_{33} \quad x_{34}]$ $\quad y_3 = 1$

$\vdots$

$x_n = [x_{n1} \quad x_{n2} \quad x_{n3} \quad x_{n4}]$ $\quad y_n = 2$

Regressione logistica

Training Data

inputs	targets / labels / ground truth	predizioni
$x_1 = [x_{11} \quad x_{12} \quad x_{13} \quad x_{14}]$	$y_1 = 1$	$\hat{y}_1 = [0.80 \quad 0.10]$
$x_2 = [x_{21} \quad x_{22} \quad x_{23} \quad x_{24}]$	$y_2 = 0$	$\hat{y}_2 = [0.30 \quad 0.70]$
$x_3 = [x_{31} \quad x_{32} \quad x_{33} \quad x_{34}]$	$y_3 = 1$	$\hat{y}_3 = [0.40 \quad 0.60]$
$\vdots$		
$x_n = [x_{n1} \quad x_{n2} \quad x_{n3} \quad x_{n4}]$	$y_n = 0$	$\hat{y}_n = [0.650 \quad 0.35]$

Regressione logistica

$x_i = [x_{i1} \quad x_{i2} \quad x_{i3} \quad x_{i4}]$ $\qquad y_i = \quad 1$ $\qquad \hat{y}_i = \quad [f_1 \quad f_2]$

$$g_1 = w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + w_4 x_{i4} + b$$

Regressione logistica

$x_i = [x_{i1} \quad x_{i2} \quad x_{i3} \quad x_{i4}]$ $\qquad y_i = \quad 1$ $\qquad \hat{y}_i = \quad [f_1 \quad f_2]$

$$logit = w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + w_4 x_{i4} + b$$

- Cosa rappresenta $\hat{y}_i$?
- Cosa è $logit$?

Regressione logistica

$x_i = [x_{i1} \quad x_{i2} \quad x_{i3} \quad x_{i4}]$ $\qquad y_i = \quad 1$ $\qquad \hat{y}_i = \quad [f_1 \quad f_2]$

$$logit = w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + w_4 x_{i4} + b$$

- Cosa rappresenta $\hat{y}_i$?
- Definiamo $f_\theta(x)$ come la funziona che fornisce le misure di probabilità per ogni classe

$$f_\theta(x) = [\Pr(y = 0|x, \theta), \Pr(y = 1|x, \theta)] \equiv [\Pr(neg|x, \theta), \Pr(pos|x, \theta)]$$

Regressione Logistica

- Preserva i classification boundaries lineari

- il decision boundary tra le classi pos e neg è determinato dalla seguente equazione:

$$\Pr(pos|x,\theta) = \Pr(neg|x,\theta)$$

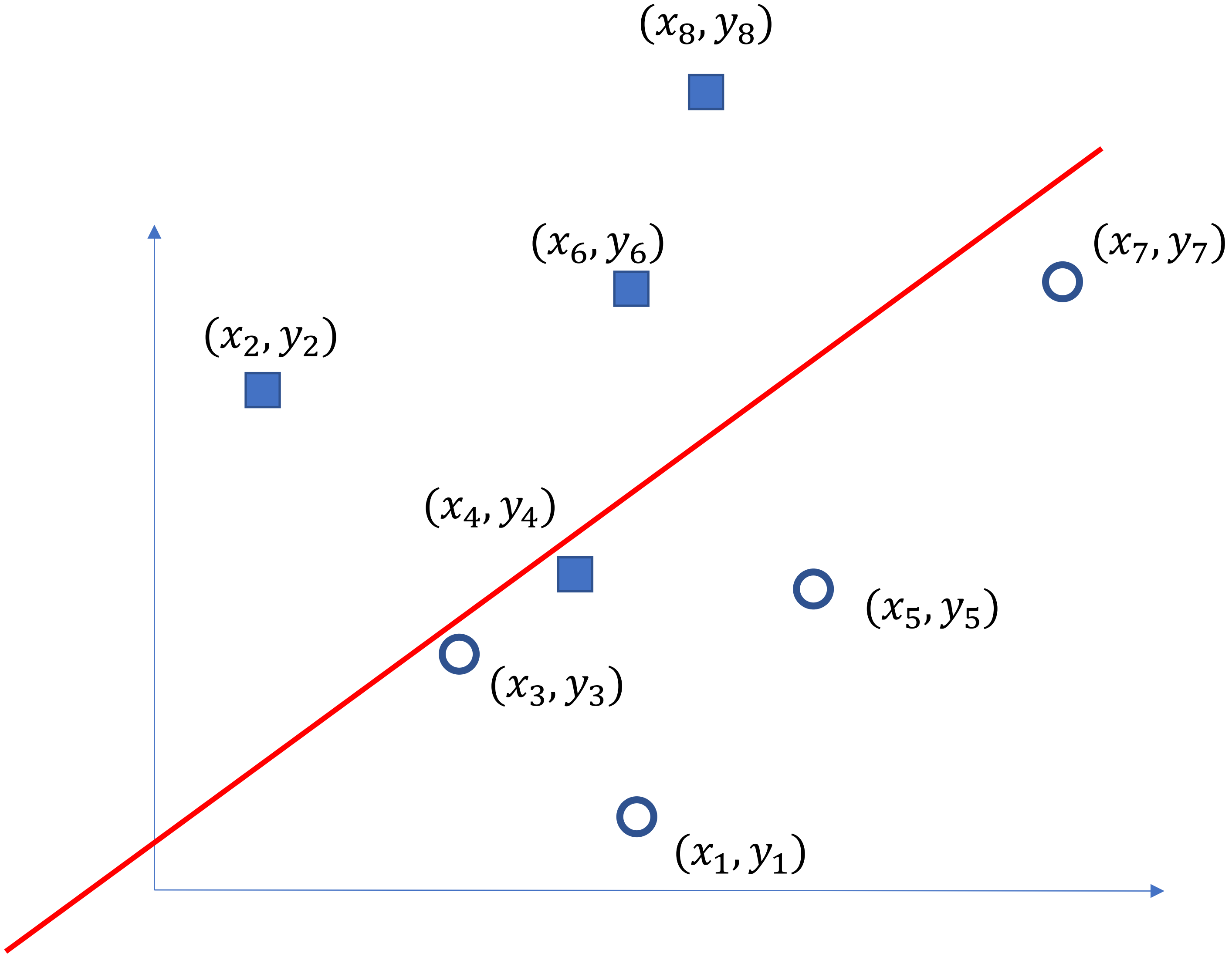
- rapportando e prendendo il logaritmo...

$$\log \frac{\Pr(pos|x,\theta)}{\Pr(neg|x,\theta)} = 0$$

- poiché il bordo deve essere lineare,

$$\log \frac{\Pr(pos|x,\theta)}{\Pr(neg|x,\theta)} = b + \sum_j w_j x_{i,j}$$

- Il rapporto è chiamato log odds, o anche logit
  - $\theta = \{b, w_1, \dots, w_m\}$ è l'insieme dei parametri da apprendere



(x_8, y_8)
(x_6, y_6)
(x_7, y_7)
(x_2, y_2)
(x_4, y_4)
(x_5, y_5)
(x_3, y_3)
(x_1, y_1)

- dalla precedente, le probabilità a posteriori diventano:

$$\Pr(pos|x,\theta) = \frac{exp\left(b + \sum_j w_j x_{i,j}\right)}{1 + exp\left(b + \sum_j w_j x_{i,j}\right)}$$

$$\Pr(neg|x,\theta) = \frac{1}{1 + exp\left(b + \sum_j w_j x_{i,j}\right)}$$

- Le probabilità sommano a 1

32

- Introduciamo le funzioni
  - Logistica

$$\sigma(x) \ = \frac{1}{1 + exp(-x)}$$

  - logit

$$\mathrm{logit}(\mathrm{x}; \theta) = b + \sum_j w_j x_{i,j}$$

- Riscriviamo le formule

$$\Pr(pos|x, \theta) \ = \sigma(logit(x; \theta))$$

$$\Pr(neg|x, \theta) = 1 - \sigma(logit(x; \theta))$$

Model learning

- Obiettivo
  - trovare i parametri $\{b, w_1, \dots, w_m\}$ che massimizzano la verosimiglianza sul training set

$$\mathcal{L}(\mathrm{D}, \theta) = \prod_{i=1}^{n} \Pr(y_i | x_i, \theta)$$

  - Equivalentemente, minimizziamo

$$nll(D, \theta) = -\sum_{i=1}^{n} \log \Pr(y_i | x_i, \theta)$$

- Se $y_i = 1$ abbiamo

$$\begin{aligned}\log \Pr(y_i|x_i,\theta) &= \log \Pr(pos|x_i,\theta)\\ &= 1 \cdot \log \Pr(pos|x_i,\theta)\\ &= y_i \cdot \log \Pr(pos|x_i,\theta)\end{aligned}$$

- analogamente, se $y_i = 0$

$$\begin{aligned}\log \Pr(y_i|x_i,\theta) &= \log(1-\Pr(pos|x_i,\theta))\\ &= 1 \cdot \log(1-\Pr(pos|x_i,\theta))\\ &= (1-y_i) \cdot \log \Pr(pos|x_i,\theta)\end{aligned}$$

- Poiché $y_i = 1$ oppure $y_i = 0$,

$$\log \Pr(y_i|x_i,\theta) = y_i \cdot \log \Pr(pos|x_i,\theta) + (1-y_i) \cdot \log(1-\Pr(pos|x_i,\theta))$$

- La verosimiglianza...

$$nll(D,\theta) = -\sum_{i=1}^{n} \log \Pr(y_i|x,\theta)$$

$$= -\sum_{i=1}^{n}\{\, y_i \cdot \log \Pr(pos|x,\theta) + \ (1-y_i)\cdot \log(1-\Pr(pos|x,\theta))\}$$

$$= -\sum_{i=1}^{n}\{\, y_i \cdot \log \hat{y}_i + \ (1-y_i)\cdot \log(1-\hat{y}_i)\}$$

$$= -\sum_{i=1}^{n} Cost(y_i,\hat{y}_i)$$

- Dove
  - $\hat{y}_i = \Pr(\mathrm{pos}|\mathrm{x},\theta)$
  - $Cost(y_i,\hat{y}_i)$= $y_i \cdot \log \hat{y}_i + \ (1-y_i)\cdot \log(1-\hat{y}_i)$

Che succede se abbiamo più classi?

Training Data

inputs

targets / labels / ground truth

$x_1 = [x_{11} \quad x_{12} \quad x_{13} \quad x_{14}]$ $\quad y_1 = 1$

$x_2 = [x_{21} \quad x_{22} \quad x_{23} \quad x_{24}]$ $\quad y_2 = 2$

$x_3 = [x_{31} \quad x_{32} \quad x_{33} \quad x_{34}]$ $\quad y_3 = 1$

$\vdots$

$x_n = [x_{n1} \quad x_{n2} \quad x_{n3} \quad x_{n4}]$ $\quad y_n = 3$

Che succede se abbiamo più classi?

Training Data

inputs	targets / labels / ground truth	Predizioni
$x_1 = [x_{11} \quad x_{12} \quad x_{13} \quad x_{14}]$	$y_1 = [1 \quad 0 \quad 0]$	$\hat{y}_1 = [0.85 \quad 0.10 \quad 0.05]$
$x_2 = [x_{21} \quad x_{22} \quad x_{23} \quad x_{24}]$	$y_2 = [0 \quad 1 \quad 0]$	$\hat{y}_2 = [0.20 \quad 0.70 \quad 0.10]$
$x_3 = [x_{31} \quad x_{32} \quad x_{33} \quad x_{34}]$	$y_3 = [1 \quad 0 \quad 0]$	$\hat{y}_3 = [0.40 \quad 0.45 \quad 0.15]$
$\vdots$		
$x_n = [x_{n1} \quad x_{n2} \quad x_{n3} \quad x_{n4}]$	$y_n = [0 \quad 0 \quad 1]$	$\hat{y}_n = [0.40 \quad 0.25 \quad 0.35]$

Che succede se abbiamo più classi?

- Un logit per ogni classe

$x_i = [x_{i1} \quad x_{i2} \quad x_{i3} \quad x_{i4}]$ $\quad y_i = [1 \quad 0 \quad 0]$ $\quad \hat{y}_i = [f_c \quad f_d \quad f_b]$

$$g_c = w_{c1}x_{i1} + w_{c2}x_{i2} + w_{c3}x_{i3} + w_{c4}x_{i4} + b_c$$

$$g_d = w_{d1}x_{i1} + w_{d2}x_{i2} + w_{d3}x_{i3} + w_{d4}x_{i4} + b_d$$

$$g_b = w_{b1}x_{i1} + w_{b2}x_{i2} + w_{b3}x_{i3} + w_{b4}x_{i4} + b_b$$

$$f_c = e^{g_c}/(e^{g_c}+e^{g_d} + e^{g_b})$$

$$f_d = e^{g_d}/(e^{g_c}+e^{g_d} + e^{g_b})$$

$$f_b = e^{g_b}/(e^{g_c}+e^{g_d} + e^{g_b})$$

Che succede se abbiamo più classi?

- La loss può essere adattata

$$nll(D, \theta) = -\sum_{i=1}^{n} Cost(y_i, \hat{y}_i)$$

$$Cost(y_i, \hat{y}_i) = \sum\nolimits_{j=1}^{k} y_{i,j} \cdot \log \hat{y}_{i,j}$$

$$\hat{y}_{i,j} = \Pr\left(y_{i,j} = 1 \middle| x_i, \theta\right) = softmax_j(x_i; \theta)$$

$$softmax_j(x_i; \theta) = \frac{\exp(logit_j(x_i, \theta))}{\sum_{h=1}^{k} \exp(logit_h(x_i, \theta))}$$

Ottimizziamo la funzione di costo

$$nll(D,\theta) = -\sum_{i=1}^{n} Cost(y_i, \hat{y}_i)$$

- Due strade:
  - Gradient-Descent (primo ordine)
  - Newton-Raphson (secondo ordine)

Gradiente discendente

$$l(\theta) \equiv nll(D,\theta) = -\sum_{i=1}^{n} Cost(y_i, \hat{y}_i)$$

- Dato λ (learning rate) e N (numero di epoche)

```
Inizializza θ_0 in maniera random
for e in range(N)
        θ_{e+1} ← θ_e − λ ∂l(θ_e)/∂θ
```

$$\theta_{e+1} \leftarrow \theta_e - \lambda \frac{\partial l(\theta_e)}{\partial \theta}$$

Idea di fondo

- Per valori opportuni di λ produce una sequenza $l(\theta_0) \geq l(\theta_1) \ldots \geq l(\theta_e)$
- Perché funziona?
  - Approssimazione in serie di Taylor al primo ordine

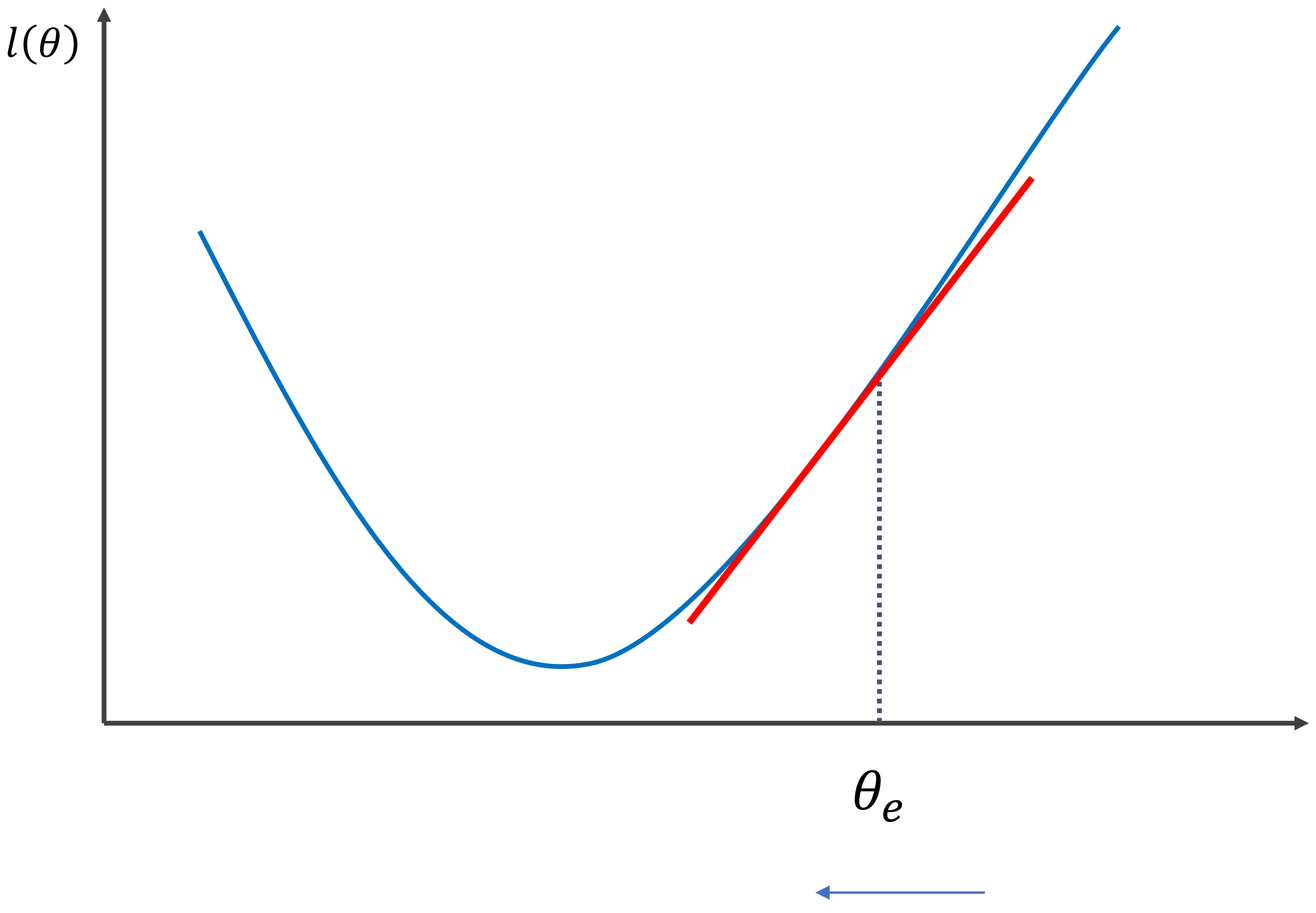
$$l(\theta) \approx l(\theta_e) + (\theta - \theta_e)^T \frac{\partial l(\theta_e)}{\partial \theta}$$

  - Calcolato sul punto θ_{e+1} e assumendo l'update

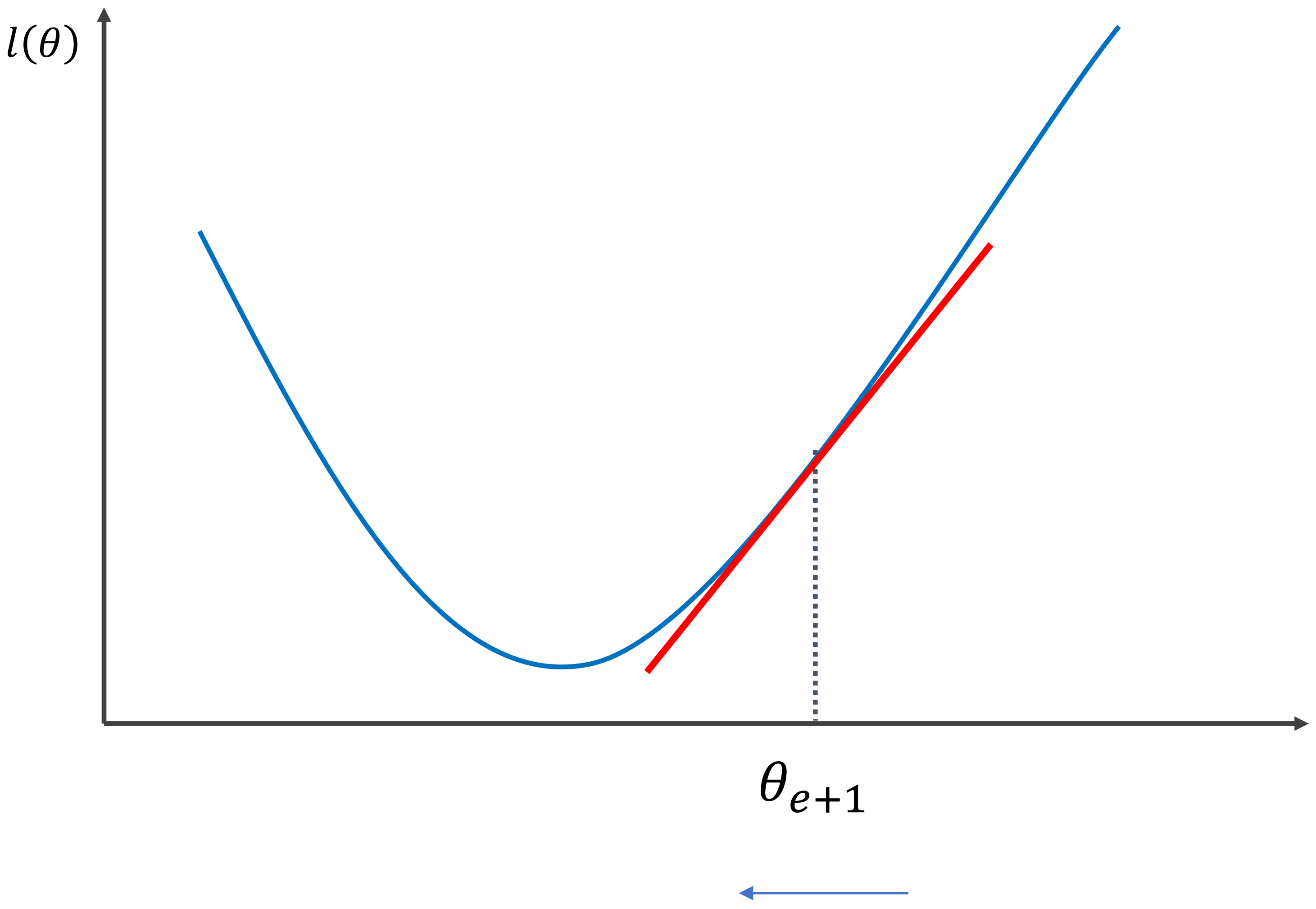
$$l(\theta_{e+1}) \approx l(\theta_e) + (\theta_{e+1} - \theta_e)^T \frac{\partial l(\theta_e)}{\partial \theta}$$
$$= l(\theta_e) - \lambda \cdot \left\| \frac{\partial l(\theta_e)}{\partial \theta} \right\|^2$$

  - Il termine $\left\| \frac{\partial l(\theta_e)}{\partial \theta} \right\|^2$ è sempre positivo

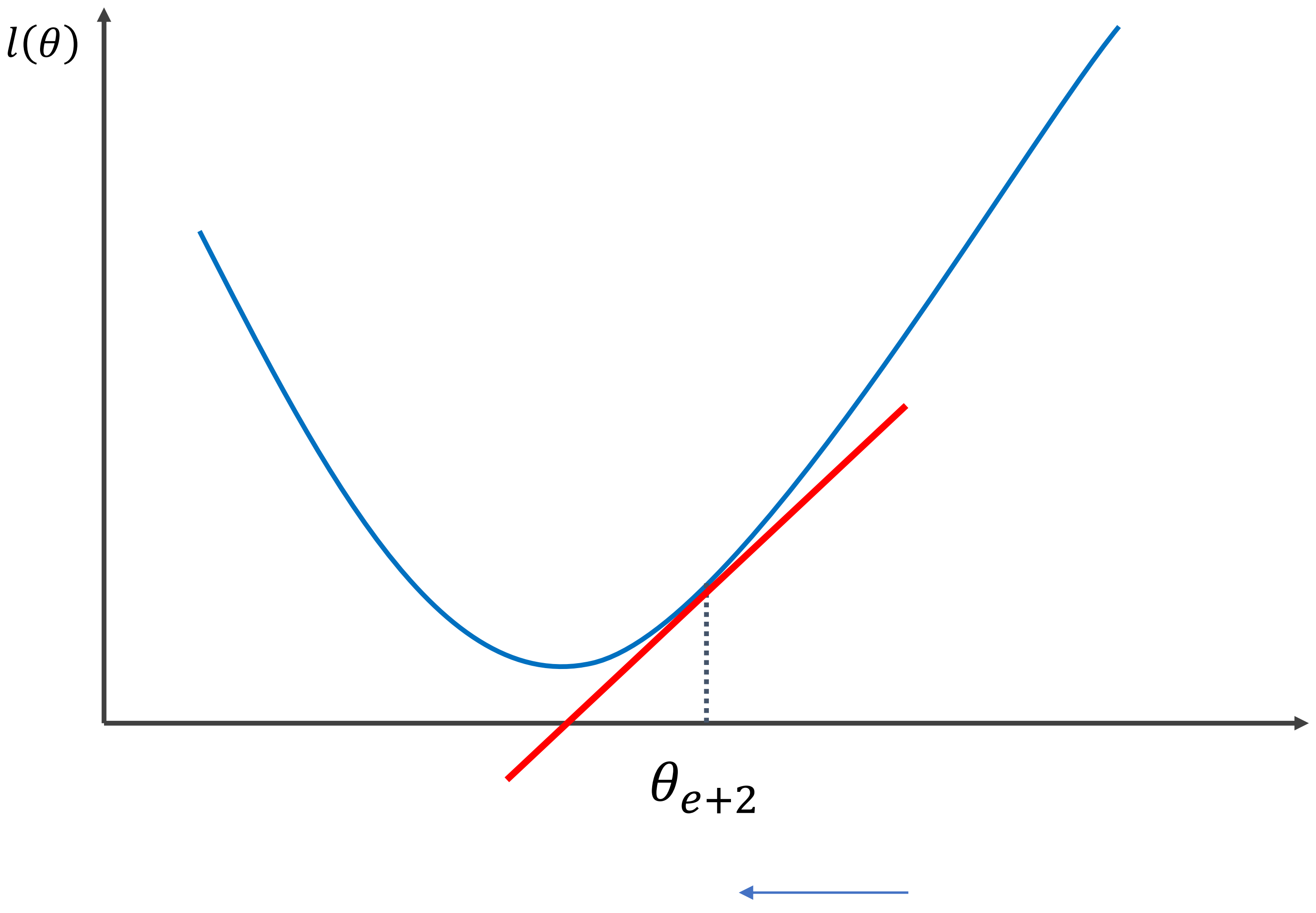
Idea di base



Idea



Idea



Gradiente discendente

costoso

$$l(\theta) \equiv nll(D,\theta) = -\sum_{i=1}^{n} Cost(y_i, \hat{y}_i)$$

- Dato λ (learning rate) e N (numero di epoche)

```
Inizializza θ0 in maniera random
for e in range(N)
        θe+1 ← θe − λ ∂l(θe)/∂θ
```

$$\theta_{e+1} \leftarrow \theta_e - \lambda \frac{\partial l(\theta_e)}{\partial \theta}$$

Minibatch (stochastic) GD

$$l_B(\theta) = -\sum_{i \in B} Cost(y_i, \hat{y}_i)$$

- Dato λ (learning rate) e N (numero di epoche), M (numero di batches)

Inizializza θ_0 in maniera random
for e in range(N)
    for b in range(M)
        $\theta_{e+1} \leftarrow \theta_e - \frac{\lambda}{B}\frac{\partial l_{B_b}(\theta_e)}{\partial\, \theta}$

Learning

- Gradient Descent

$$\theta_{e+1} \leftarrow \theta_e \ - \lambda \frac{\partial l\ (\theta_e)}{\partial\ \theta}$$

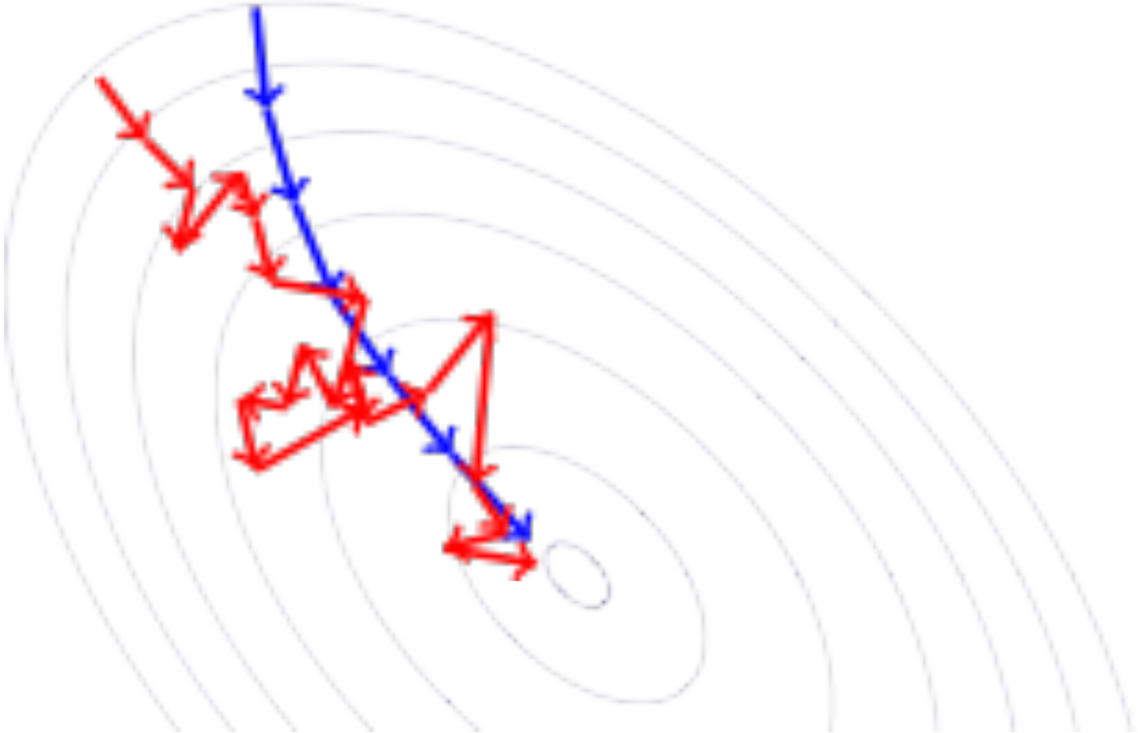
- Stochastic GD

$$\theta_{e+1} \leftarrow \theta_e \ - \frac{\lambda}{N} \frac{\partial l_i(\theta_e)}{\partial\ \theta}$$

- Mini-batch SGD

$$\theta_{e+1} \leftarrow \theta_e \ - \frac{\lambda}{B} \sum_{i \in B} \frac{\partial l_i(\theta_e)}{\partial\ \theta}$$

$$l_i(\theta) = Cost(y_i, \hat{y}_i)$$



More to come later …

- Regularization
- Momentum updates
- Hinge Loss, Least Squares Loss, Logistic Regression Loss…

Riassumendo

$$\hat{y}_i = [f_c \quad f_d \quad f_b]$$

Estraiamo le features

$$x_i = [x_{i1} \quad x_{i2} \quad x_{i3} \quad x_{i4}]$$

Calcoliamo I logits

$$g_c = w_{c1}x_{i1} + w_{c2}x_{i2} + w_{c3}x_{i3} + w_{c4}x_{i4} + b_c$$

$$g_d = w_{d1}x_{i1} + w_{d2}x_{i2} + w_{d3}x_{i3} + w_{d4}x_{i4} + b_d$$

$$g_b = w_{b1}x_{i1} + w_{b2}x_{i2} + w_{b3}x_{i3} + w_{b4}x_{i4} + b_b$$

$$f_c = e^{g_c}/(e^{g_c}+e^{g_d} + e^{g_b})$$

$$f_d = e^{g_d}/(e^{g_c}+e^{g_d} + e^{g_b})$$

$$f_b = e^{g_b}/(e^{g_c}+e^{g_d} + e^{g_b})$$

Otteniamo le predizioni