

# Fondamenti di Machine Learning

Overfitting e Tecniche di Regolarizzazione

## Tradeoff Bias/Varianza

presentato da  
Giuseppe Gullo

# Bias e Varianza

Sono due concetti legati all'errore del modello  
e inversamente proporzionati tra di loro.

# Bias e Varianza

Sono due concetti legati all'errore del modello  
e inversamente proporzionati tra di loro.

Il **bias** è la differenza tra la previsione media e il valore reale.

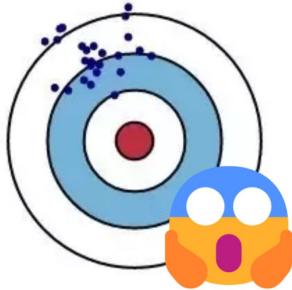
# Bias e Varianza

Sono due concetti legati all'errore del modello  
e inversamente proporzionati tra di loro.

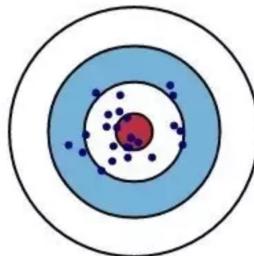
Il **bias** è la differenza tra la previsione e valore reale.

La **varianza** è la variabilità del modello  
per un determinato dato.

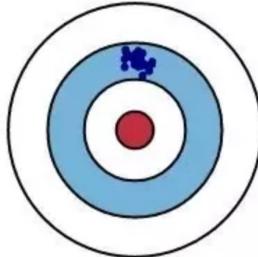
**ALTO BIAS  
ALTA VARIANZA**



**BASSO BIAS  
ALTA VARIANZA**



**ALTO BIAS  
BASSA VARIANZA**

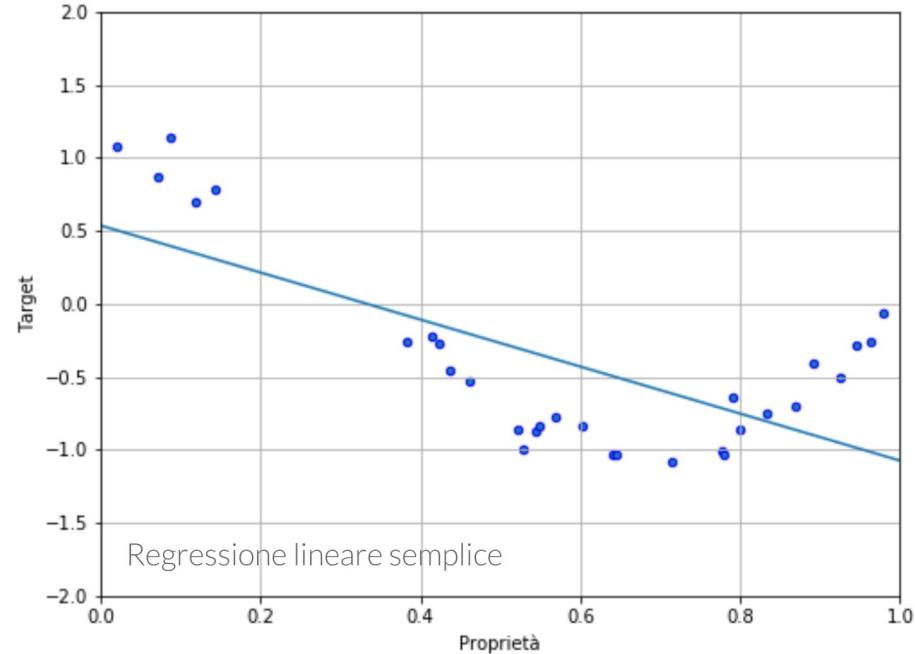


**BASSO BIAS  
BASSA VARIANZA**



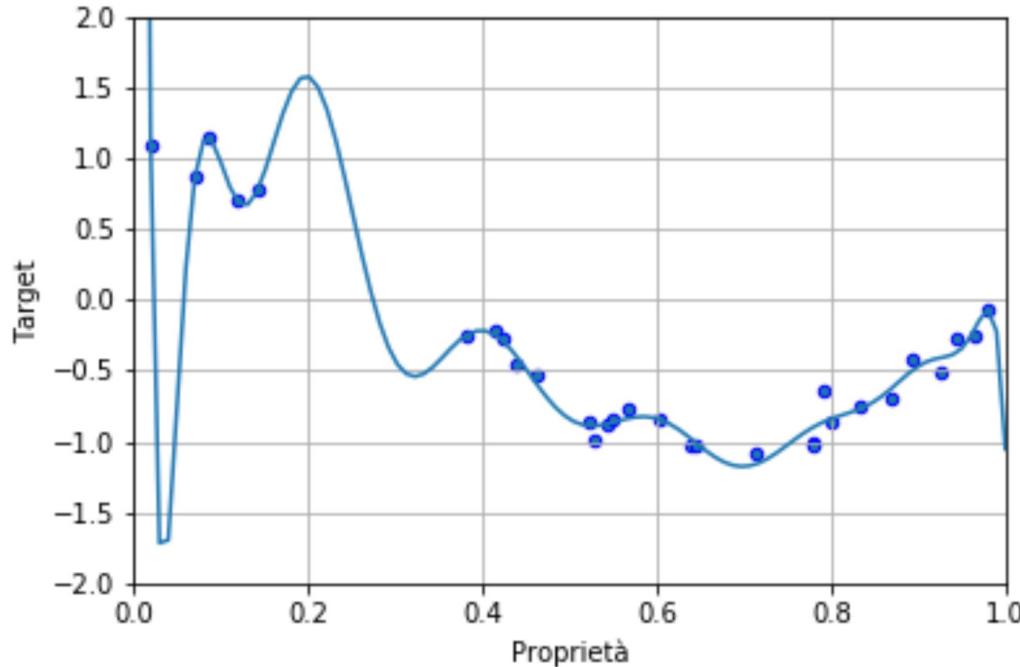
## Alto Bias Bassa Varianza

In questo caso si parla di **underfitting**, il modello è troppo semplice



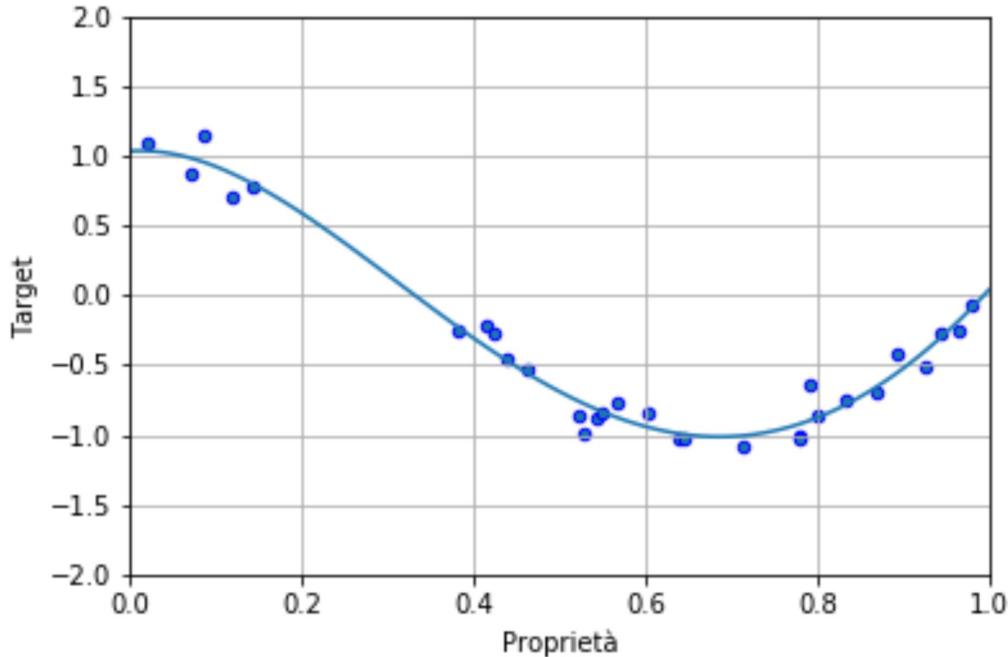
## Alta varianza Basso Bias

In questo caso si parla di **overfitting**, il modello è troppo complesso



## Compromesso Bias / Varianza

Il modello migliore si realizza bilanciando bias e varianza

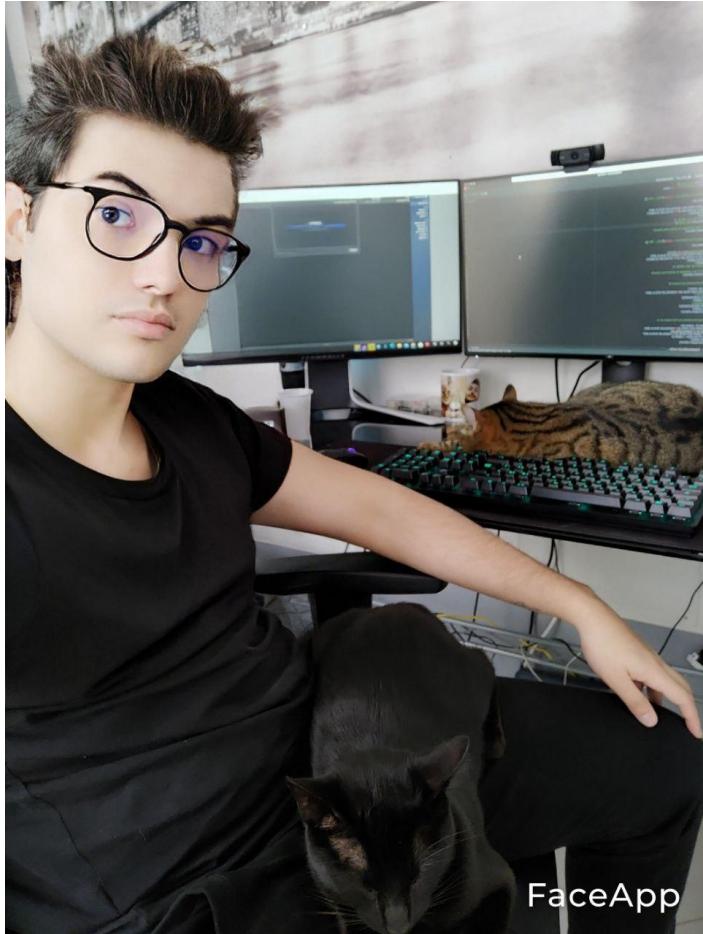


# Fondamenti di Machine Learning

Overfitting e Tecniche di Regolarizzazione

## Capire l'overfitting

presentato da  
Giuseppe Gullo



# Lui è Giuseppino

# Questo è un quadrato



# Questo è un quadrato

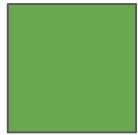
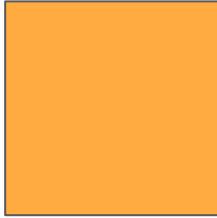


# Questo è un quadrato



$$Area = base \cdot altezza$$

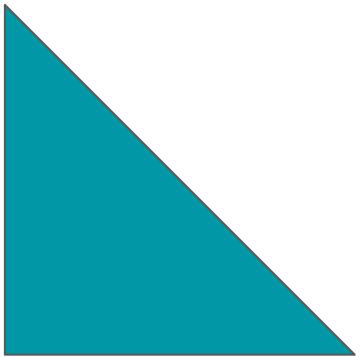
# Questi sono altri quadrati



$$Area = base \cdot altezza$$

10 e lode!

# Questo è un triangolo rettangolo



*Area = ?*

# Questo è un triangolo rettangolo



*Area = ?*

# Questo è un triangolo rettangolo



$$Area = \frac{base \cdot altezza}{2}$$

# Questo è l'overfitting

Perché Giuseppino non ha compreso la regola,  
la ha solo memorizzata e applicata.

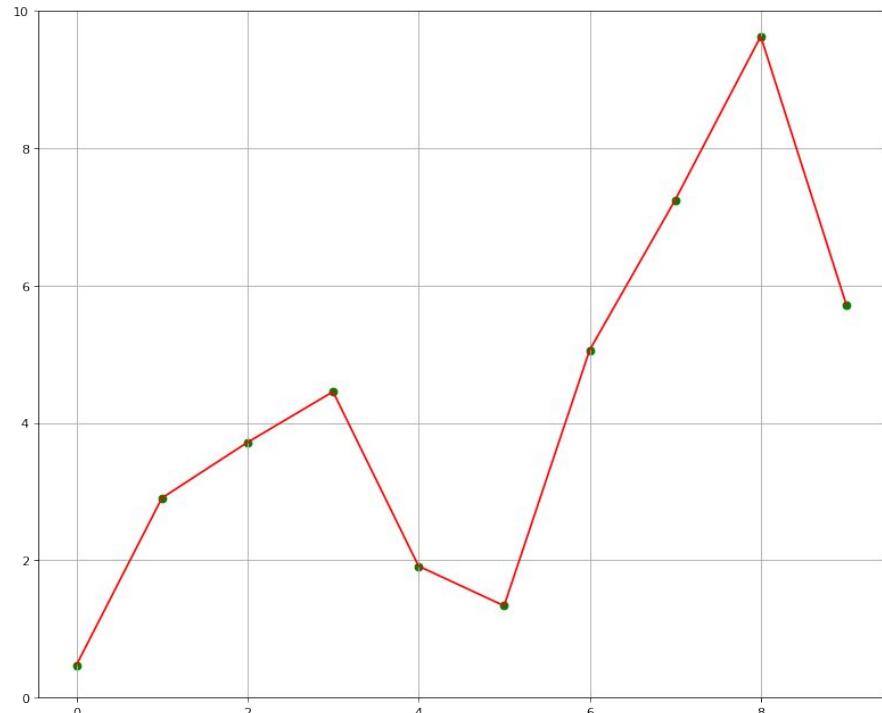
# Questo è l'overfitting

Perché Giuseppino non ha compreso la regola,  
la ha solo memorizzata e applicata.

Quindi se gli si presenta un problema lievemente differente  
non è in grado di risolverlo.

# Overfitting nel Machine Learning

Il modello non apprende dai dati, li memorizza.



# Overfitting nel Machine Learning

Un modello che soffre di overfitting  
tende a sovrastimare le sue performance reali  
non riesce a generalizzare su esempi non noti.

# Fondamenti di Machine Learning

Overfitting e Tecniche di Regolarizzazione

## Hold-out vs e Cross-validation

presentato da  
Giuseppe Gullo

# Come si riconosce l'overfitting?

Confrontando le performance sul set di addestramento  
e su un set di dati sconosciuti al modello.

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	170.000
60	150.000

## Hold-out

Dividiamo in dataset in due porzioni,  
una viene usata per l'addestramento l'altra per la valutazione

## Hold-out

Dividiamo in dataset in due porzioni,  
una viene usata per l'addestramento l'altra per la valutazione

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	170.000
60	150.000

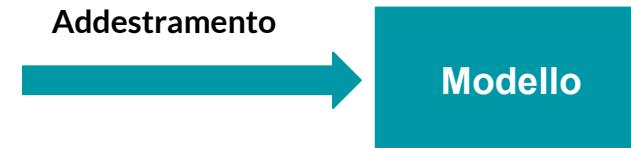
# Addestramento

## Train set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000

## Test set

Metri quadri	Valore in €
110	250.000
70	170.000
60	150.000



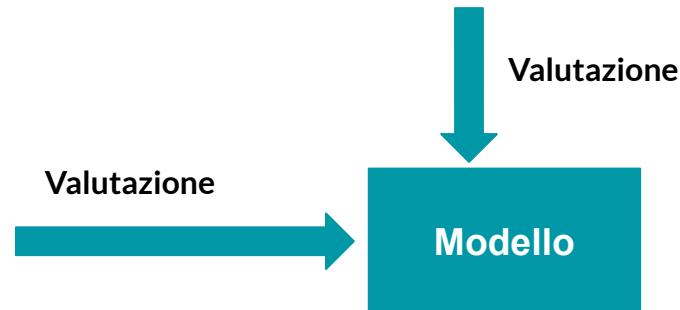
# Valutazione

## Train set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000

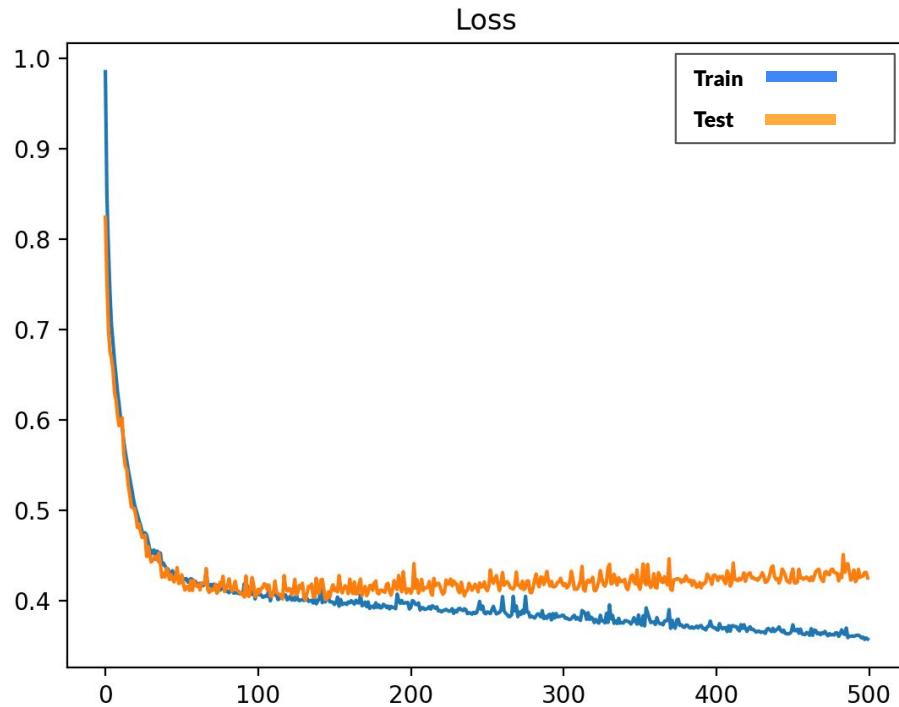
## Test set

Metri quadri	Valore in €
110	250.000
70	170.000
60	150.000



# La learning curve

Ci permette di confrontare l'andamento dell'addestramento all'aumentare del numero di esempi o di epocha.



Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	170.000
60	150.000

## Cross-validation

Dividiamo il dataset in K porzioni, per ogni K,  
addestriamo il modello su K-1 porzioni  
e testiamo su 1 porzione

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000

Metri quadri	Valore in €
110	200.000
110	250.000
70	170.000
60	150.000

Metri quadri	Valore in €
50	100.000
120	240.000
60	180.000

**K = 3**

# Addestramento

## Train set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
Metri quadri	Valore in €
50	100.000
120	240.000
60	180.000

## Test set

Metri quadri	Valore in €
110	200.000
110	250.000
70	170.000
60	150.000

Addestramento



Hold-out e Cross-validation

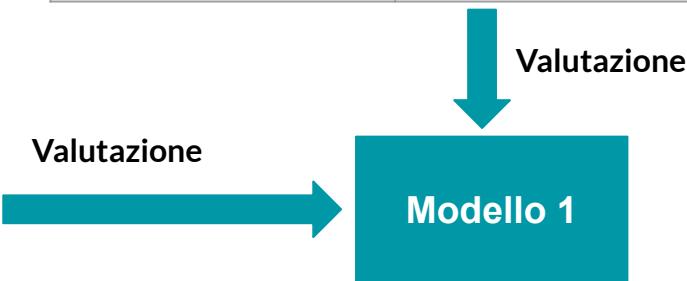
# Addestramento

## Train set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
Metri quadri	Valore in €
50	100.000
120	240.000
60	180.000

## Test set

Metri quadri	Valore in €
110	200.000
110	250.000
70	170.000
60	150.000



# Addestramento

## Train set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
Metri quadri	Valore in €
110	200.000
110	250.000
70	170.000
60	150.000

## Test set

Metri quadri	Valore in €
50	100.000
120	240.000
60	180.000

Addestramento



Modello 2

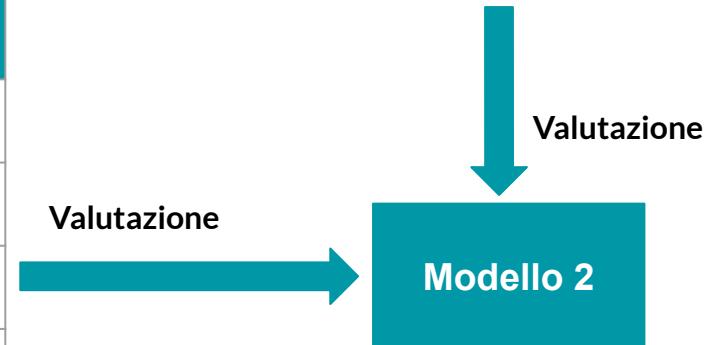
# Addestramento

## Train set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
Metri quadri	Valore in €
110	200.000
110	250.000
70	170.000
60	150.000

## Test set

Metri quadri	Valore in €
50	100.000
120	240.000
60	180.000



# Addestramento

## Train set

Metri quadri	Valore in €
50	100.000
120	240.000
60	180.000
Metri quadri	Valore in €
110	200.000
110	250.000
70	170.000
60	150.000

## Test set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000

Addestramento



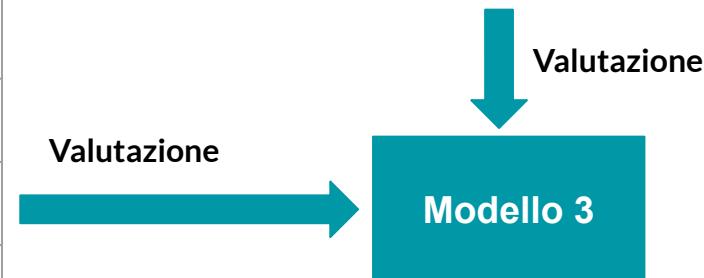
# Addestramento

## Train set

Metri quadri	Valore in €
50	100.000
120	240.000
60	180.000
Metri quadri	Valore in €
110	200.000
110	250.000
70	170.000
60	150.000

## Test set

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000



# Cross-validation

Possiamo fare la media delle performance sui 3 modelli

Modello 1

Modello 2

Modello 3

# Cross-validation

Possiamo fare la media delle performance sui 3 modelli

Modello 1

Modello 2

Modello 3

Dopo la valutazione, il modello va addestrato sull'intero dataset

# Train - Validation - Test

Se eseguiamo un numero elevato di training e testing del modello,  
il modello rischia di andare in overfitting sul set di test.

# Train - Validation - Test

Se eseguiamo un numero elevato di training e testing del modello,  
il modello rischia di andare in overfitting sul set di test.

Questo è comune nella fasi di **tuning degli iperparametri**.

# Train - Validation - Test

Se eseguiamo un numero elevato di training e testing del modello, il modello rischia di andare in overfitting sul set di test.

Questo è comune nelle fasi di **tuning degli iperparametri**.

- Utilizziamo il **validation set** per confrontare diverse configurazioni di uno stesso modello.
- Utilizziamo il **test set** per confrontare diversi modelli.

# Train - Validation - Test

Se eseguiamo un numero elevato di training e testing del modello, il modello rischia di andare in overfitting sul set di test.

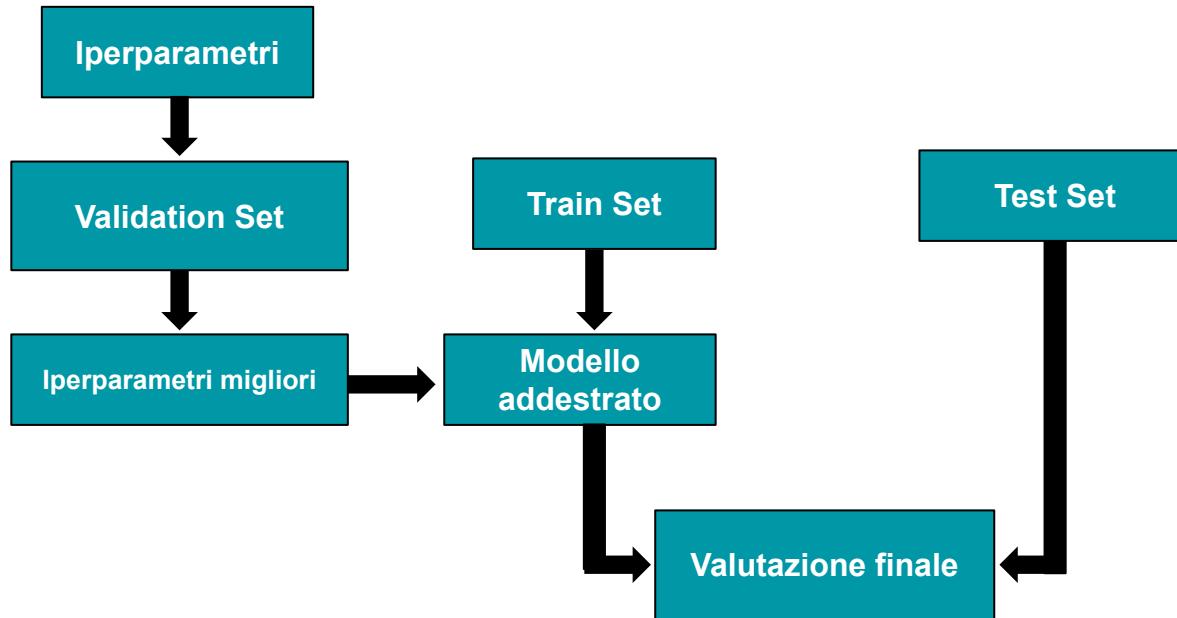
Questo è comune nelle fasi di **tuning degli iperparametri**.

- Utilizziamo il **validation set** per confrontare diverse configurazioni di uno stesso modello.
- Utilizziamo il **test set** per confrontare diversi modelli.

Spesso validation set e test set vengono utilizzati come sinonimi

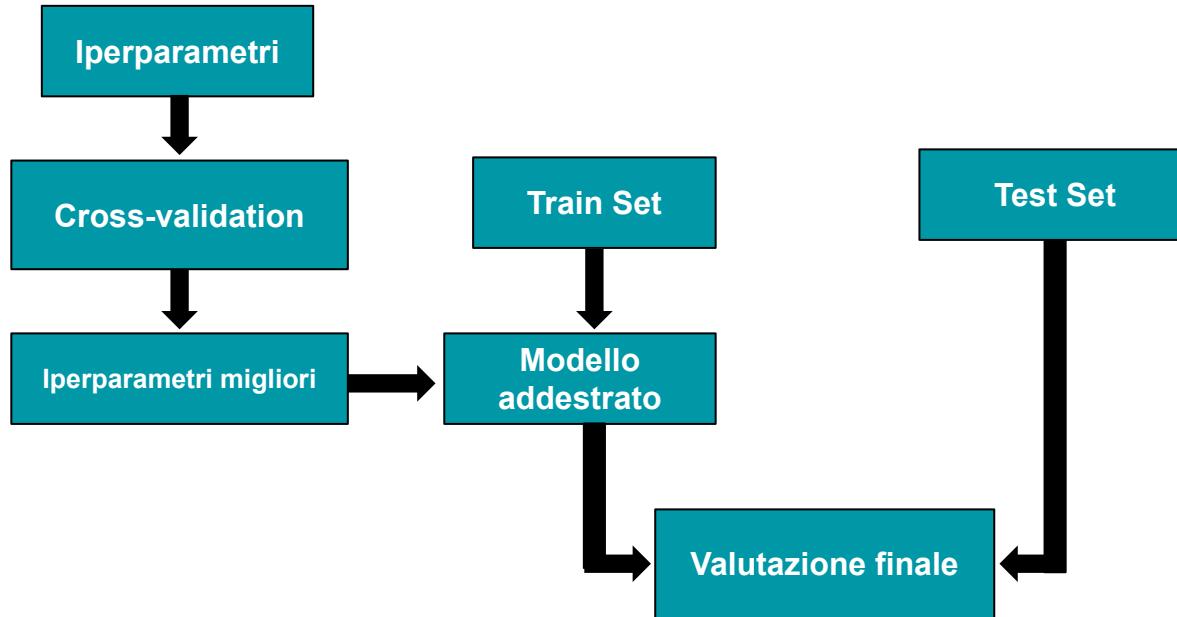
# Train - Validation - Test

Utilizziamo un set di validazione per ottimizzare gli iperparametri



# Train - Validation - Test

Possiamo utilizzare la cross-validation al posto del set di validazione



# Hold-out o Cross-validation?

- La cross-validation è un indicatore migliore, ma è molto dispendiosa.
- L'Hold-out è meno indicativo, ma è anche meno dispendioso.

# Come affrontare l'overfitting?

1. Raccogliere più dati
2. Rimuovere feature per ridurre la complessità del modello
3. Applicare tecniche di regolarizzazione

# Come affrontare l'overfitting?

1. Raccogliere più dati
2. Rimuovere feature per ridurre la complessità del modello
3. **Applicare tecniche di regolarizzazione**

# Fondamenti di Machine Learning

Overfitting e Tecniche di Regolarizzazione

## La Regolarizzazione L1 e L2

presentato da  
Giuseppe Gullo

## Regolarizzazione L1 e L2

Permettono di contrastare l'overfitting  
penalizzando i pesi maggiori

$W \gg w$

## Regolarizzazione L2 - Weight Decay

$$L(w, b) = \sum_{i=1}^N (y_i - wx + b)^2 + \lambda \sum_{j=1}^M w_j^2$$

## Regolarizzazione L2 - Weight Decay

$$L(w, b) = \sum_{i=1}^N (y_i - wx + b)^2 + \lambda \sum_{j=1}^M w_j^2$$

$\lambda$  è il **parametro di regolarizzazione**  
ed è un iperparametro del modello.

## Regolarizzazione L2 - Weight Decay

$$L(w, b) = \sum_{i=1}^N (y_i - wx + b)^2 + \lambda \sum_{j=1}^M w_j^2$$

Nel caso di regolarizzazione L2 si parla di **Regressione Ridge**

# Regolarizzazione L1

$$L(w, b) = \sum_{i=1}^N (y_i - wx + b)^2 + \lambda \sum_{j=1}^M |w_j|$$

# Regolarizzazione L1

$$L(w, b) = \sum_{i=1}^N (y_i - wx + b)^2 + \lambda \sum_{j=1}^M |w_j|$$

Nel caso di regolarizzazione L1 si parla di **Regressione Lasso**

## Qual è la differenza?

La regolarizzazione L1 è molto più intensa della L2  
e tende a portare molti pesi a 0.

$$W = [w_1, w_2, w_3, w_4, w_5]$$



$$W = [w_1, 0, 0, w_4, 0]$$

## Qual è la differenza?

La regolarizzazione L1 è molto più intensa della L2  
e tende a portare molti pesi a 0.

$$W = [w_1, w_2, w_3, w_4, w_5]$$



$$W = [w_1, 0, 0, w_4, 0]$$

Ci permette di fare feature selection

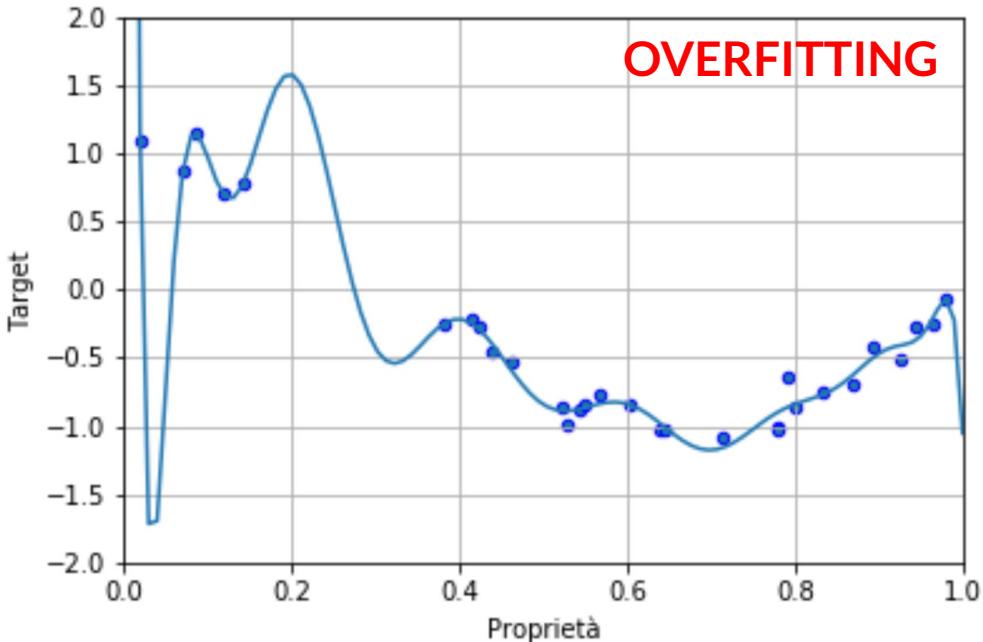
# Come scegliere il parametro di regolarizzazione?

Cercare in uno spazio di potenze di 10

$$10^{-4} \leq \lambda \leq 10$$

# Come scegliere il parametro di regolarizzazione?

$\lambda = 0$  Nessuna regolarizzazione



OVERFITTING

# Come scegliere il parametro di regolarizzazione?

$$\lambda = \infty$$

Tutti i pesi vengono portati a 0

