

1/50



Reti neurali artificiali per lo studio di relazioni non lineari tra fattori e item

Confronto tra PCA e autoencoder non lineare



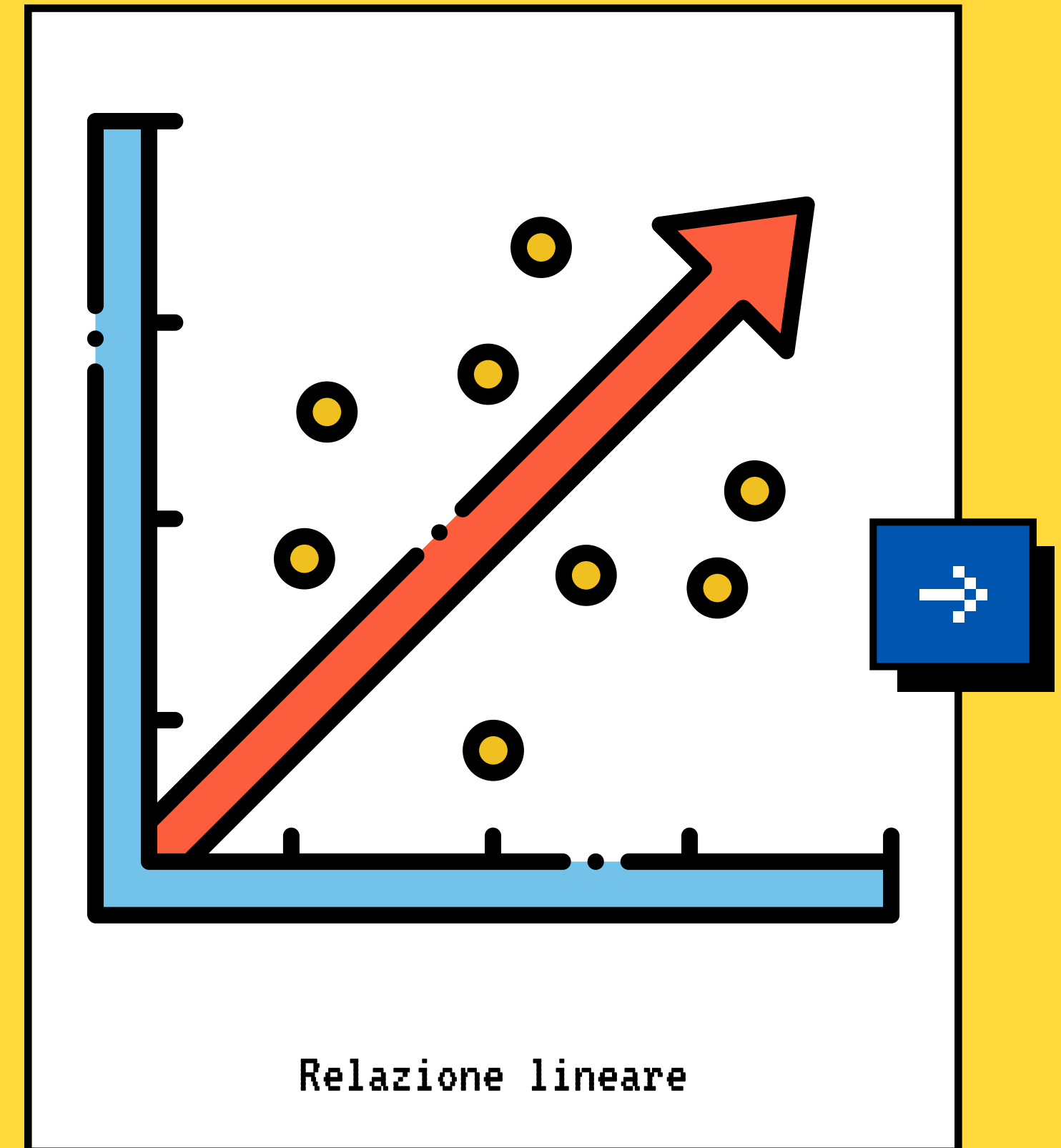
I modelli esplicativi lineari



I modelli esplicativi presuppongono che gli effetti esercitati dalle variabili indipendenti sulle variabili dipendenti siano lineari



La forza della relazione tra un fattore e una variabile misurata è costante almeno all'interno dell'intervallo osservato dei dati



Violazioni dell'assunzione di linearità

✓ Effetto pavimento

il livello minimo al di sotto del quale la varianza di una variabile indipendente non è più misurabile

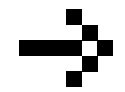
✓ Effetto soffitto

il livello massimo al di sopra del quale la varianza di una variabile indipendente non è più misurabile

✓ Inversioni di direzione

Livelli crescenti del fattore possono produrre differenze sia decrescenti sia crescenti nella variabile misurata

Esempi



Aggressività

Risulta più difficile giudicare le differenze tra bassi livelli di aggressività rispetto ad alti livelli di aggressività

Atteggiamento

I rispondenti si collocano agli estremi più spesso di quanto ci si aspetterebbe sulla base di un modello lineare.

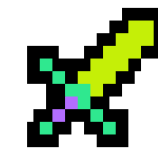
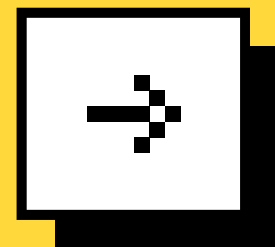
Variabili qualitative

Trattare variabili dicotomiche o ordinali come continue può produrre una non linearità nella relazione tra item e fattori.

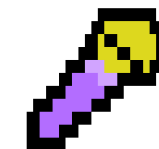
4/50



Conseguenze



La conseguenza dell'assunzione della linearità è sostanzialmente l'eccessiva semplificazione della realtà, se non addirittura l'errata specificazione dei modelli.



Ciò può comportare, ad esempio:



Il rifiuto dell'invarianza di misura di un costrutto;

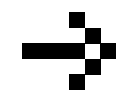


La confusione tra effetti non lineari ed effetti di interazione.



Escludere la non linearità dai modelli psicometrici limita le possibilità di comprensione dei fenomeni psicologici.

L'invarianza di misura



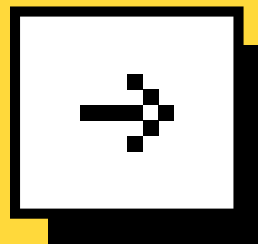
Equivalenza del significato del costrutto tra i gruppi o nel tempo

- ✚ Equivalenza **psicometrica** di un costrutto o di un insieme di costrutti in due o più gruppi o nel tempo
- ✚ Confronto di campioni indipendenti
- ✚ Analisi fattoriale confermativa a gruppi multipli per testare la sostenibilità dei vincoli di uguaglianza sulle varie componenti del modello di misurazione.

6/50



Tipi di invarianza



Invarianza configurale

Equivalenza delle forme dei modelli

Invarianza metrica

Equivalenza delle saturazioni dei modelli

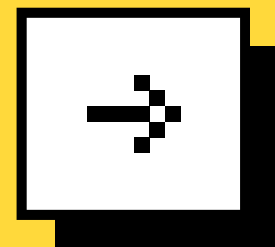
Invarianza scalare

Equivalenza delle intercette dei modelli

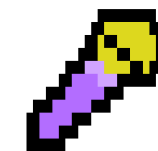
Invarianza residua

Equivalenza dei residui dei modelli

La linearità



E IL RIFIUTO DEI
TEST DI INVARIANZA
DI MISURA



Il modello di analisi fattoriale utilizzato per valutare l'invarianza di misura presuppone che le relazioni delle osservazioni con i fattori siano lineari.



Quando l'invarianza di misura è rifiutata in un modello fattoriale lineare, la ragione potrebbe essere che la vera relazione tra fattore e indicatore è non lineare.

Bauer (2005)

Relazione quadratica:

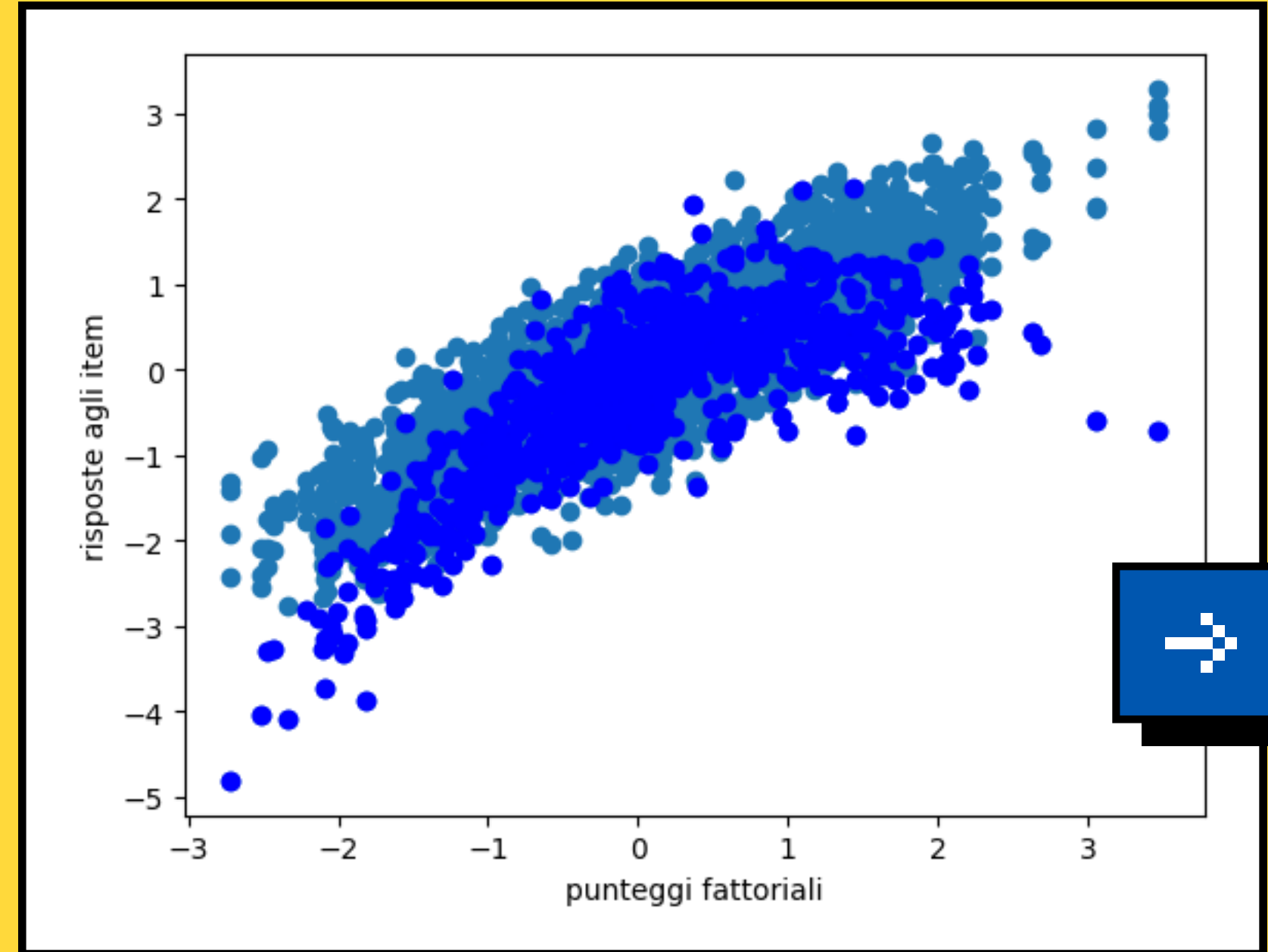
$$y = ax^2 + bx + c$$

Relazione lineare:

$$y = bx + c$$

$$b = \frac{COV_{xy}}{VAR_{xy}}$$

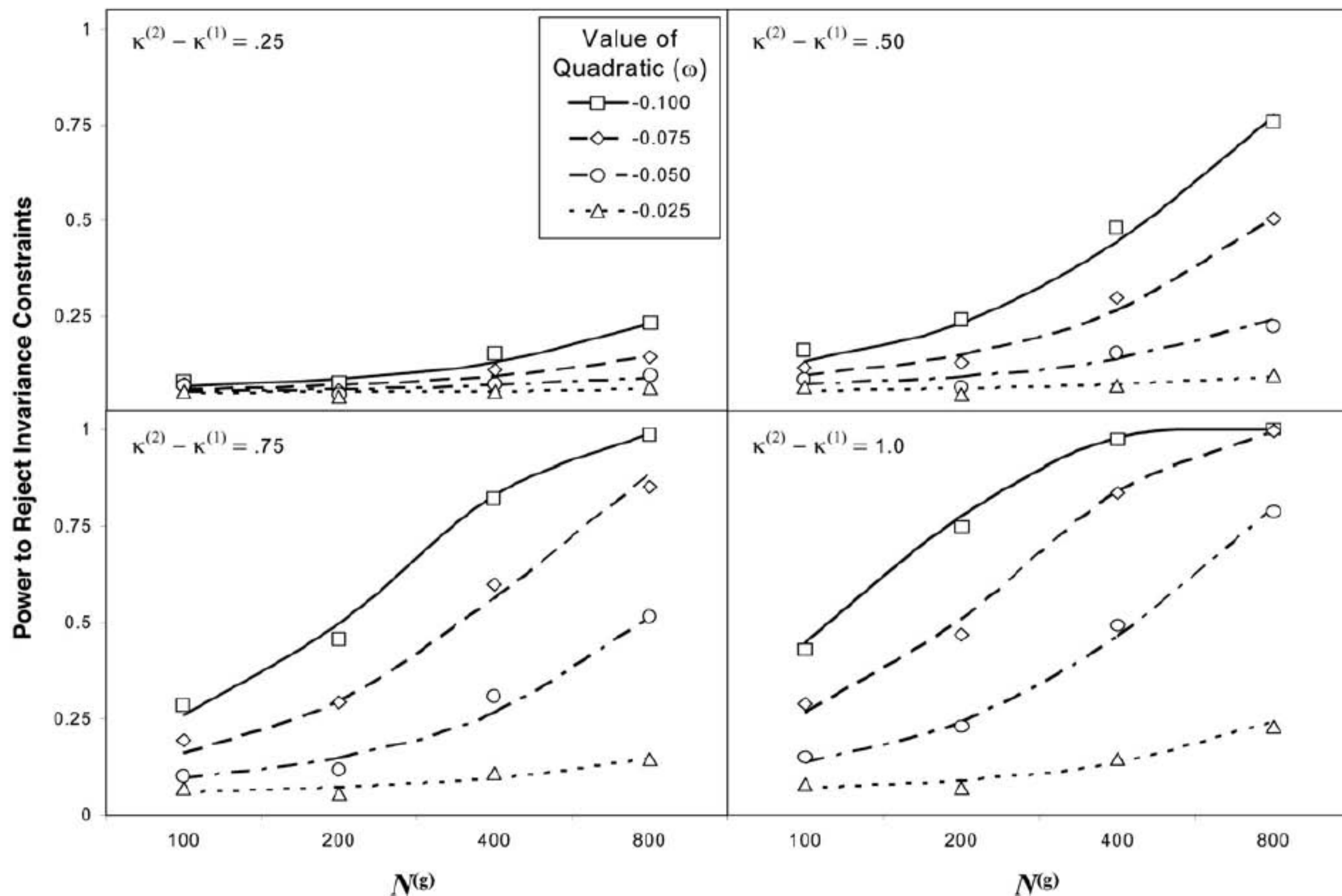
$$c = \bar{Y} - (b\bar{X})$$




Dataset


4 item lineari


1 item quadratico



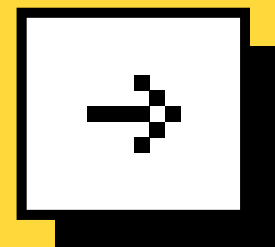
 **Rifiuto dei test di invarianza metrica e scalare, specialmente con l'aumento di:**

 **N**

 **Differenza delle medie fattoriali**

 **Diverse pendenze e intercette tra i due gruppi**

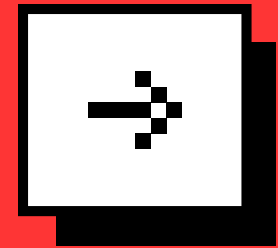
Implicazioni




In generale, concludere che c'è un'invarianza **parziale** può portare a conclusioni sostanzialmente errate se la relazione è davvero invariante ma non lineare.

Data la forte dipendenza di questi risultati dalle medie dei fattori, questo problema si presenterà molto probabilmente in quelle applicazioni in cui si prevede una grande differenza di media dei fattori (confronto tra gruppi di età, punteggi pre e post-test, gruppi etnici valutati con misure culturalmente sensibili).

Effetti di interazione ed effetti non lineari



 Un effetto di interazione comporta che l'effetto di un predittore su un fenomeno dipende dal valore di un altro predittore.

La mancata considerazione di possibili effetti non lineari può portare alla falsa individuazione di effetti di interazione.

Busemeyer e Jones (1983): non tenere conto dell'effetto quadratico di un predittore può portare all'identificazione di un falso effetto di interazione per questo predittore con un altro predittore correlato.

Una percentuale sconosciuta degli effetti di interazione in letteratura potrebbe essere spuria e riflettere l'omissione di effetti non lineari dai modelli adattati.

12/50



Effetti di interazione ed effetti non lineari



In un'equazione di regressione, un effetto quadratico può essere mascherato da un effetto di interazione (o viceversa) a causa della loro somiglianza matematica.



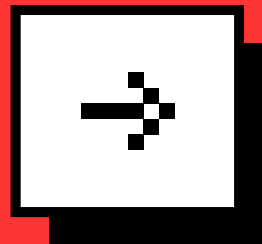
Effetto non lineare

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot Z_i + \beta_3(X_i \cdot X_i) + \varepsilon_i$$



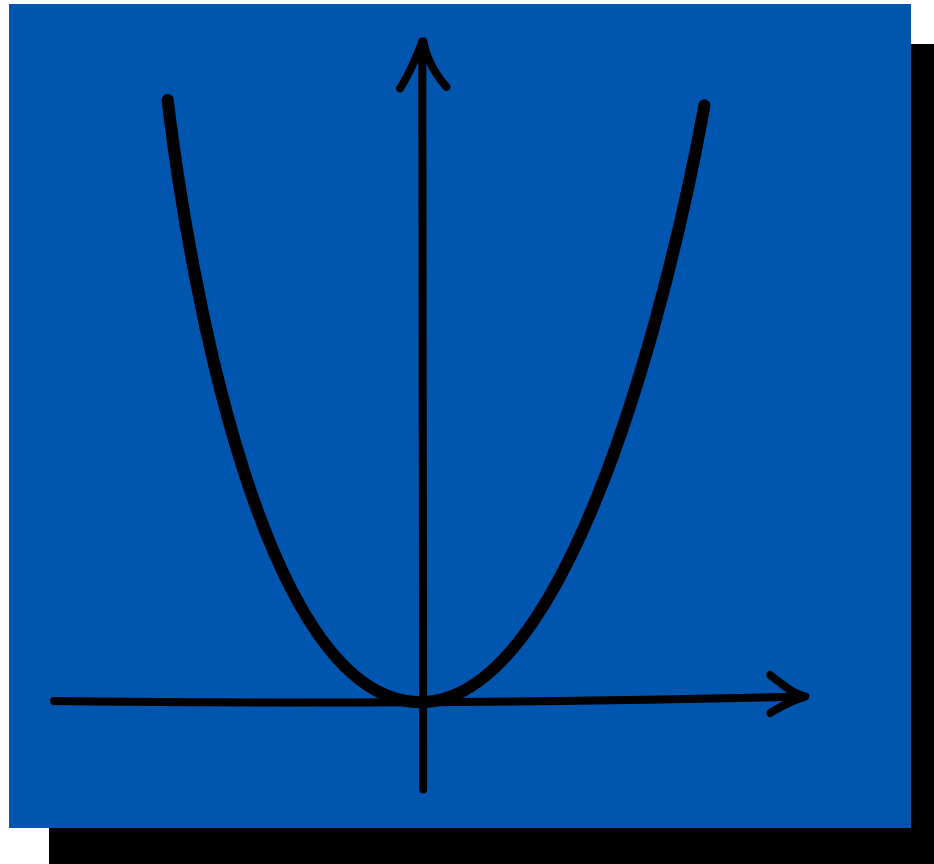
Effetto di interazione

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot Z_i + \beta_3(X_i \cdot Z_i) + \varepsilon_i$$

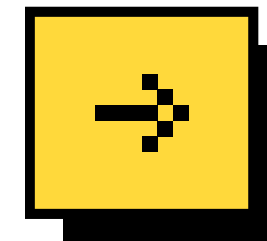


Falsa identificazione di un effetto di interazione quando in realtà non è presente

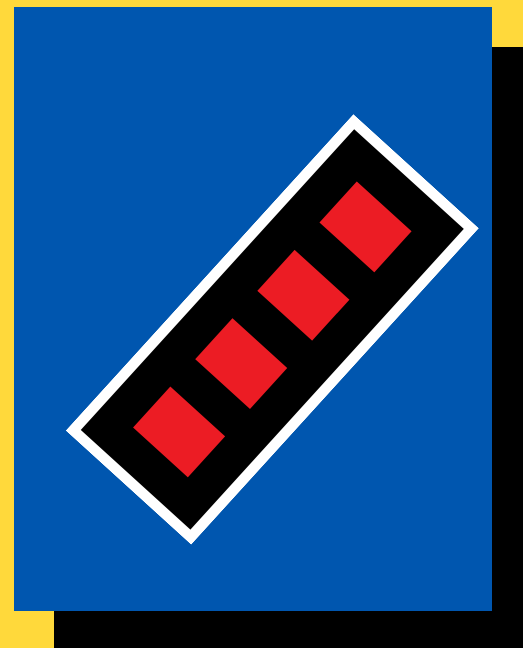
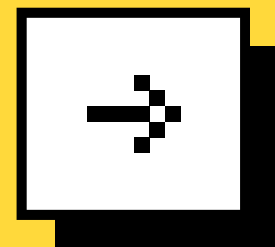
Modelli non lineari



I metodi per stimare modelli non lineari sono raramente implementati, nonostante siano disponibili da molti anni insieme a software specializzati. Questa negligenza può riflettere l'idea che piccole deviazioni dalla linearità siano relativamente innocue, in quanto un modello lineare fornirà comunque un'utile approssimazione di primo ordine alla funzione vera.



Modelli esplicativi non lineari



Primi contributi:

01

- McDonald (1967)
- Etezadi-Amoli e McDonald (1983)
- Mooijaart e Bentler (1986),

Successivamente, sono stati
sviluppati nuovi metodi di
stima:

02

- basati sul **maximum-likelihood** [Klein e Moosbrugger (2000), Klein e Muthe ´n (2003), Lee e Zhu (2002) e Yalcin e Amemiya (2001)];
- basati sui **momenti** [Wall e Amemiya (2000)];
- **bayesiani** [Arminger e Muthe ´n (1998) e Zhu e Lee (1999)]

Modelli descrittivi non lineari

Si è cominciato a prestare attenzione ai dati con struttura non lineare a partire dal 2000, quando sono usciti ISOMAP (Isometric Mapping) e LLE (Locally Linear Embedding).

✓ Metodi globali e locali

Isomap è considerato un metodo globale, in quanto ricava un sottospazio a più bassa dimensionalità dalla distanza geodetica tra tutte le coppie di punti. LLE è considerato un metodo locale perché la funzione di costo che utilizza per costruire il sottospazio considera solo il posizionamento di ogni punto rispetto ai suoi vicini.

✓ Tecniche convesse e non convesse

Isomap e LLE sono algoritmi di apprendimento spettrale, perché per ottenere la rappresentazione dei dati nello spazio a più bassa dimensionalità si utilizza la decomposizione spettrale (o decomposizione agli autovalori). Le tecniche spettrali sono tecniche convesse, nelle quali la funzione obiettivo non ha alcun optima locale, ma esistono anche tecniche non convesse.

Problemi dei modelli non lineari esplicativi e descrittivi

01

Sono computazionalmente
costosi

02

La loro caratteristica di operare
ipotesi predefinite sulla forma
matematica delle relazioni tra le
variabili può renderli poco adatti a
cogliere relazioni non lineari, nei
casi in cui le relazioni tra le
variabili sono complesse

Autoencoder per la riduzione della dimensionalità

Obiettivo

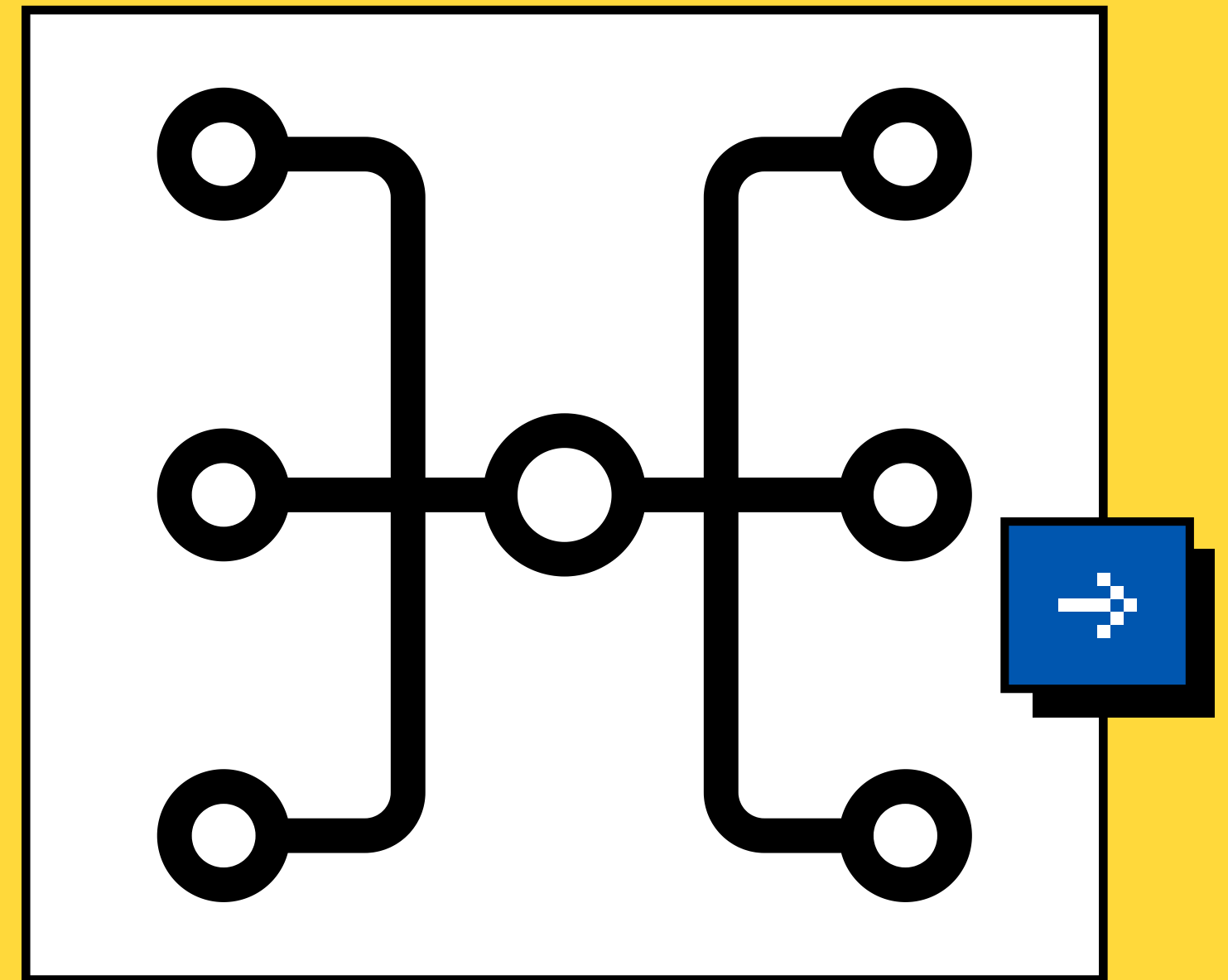
01

Introdurre un metodo predittivo, basato sulle reti neurali artificiali, per ridurre la dimensionalità di dataset che presentano relazioni non lineari tra item e fattori.

02

Ipotesi

Un autoencoder non lineare avrebbe superato la PCA nel ricostruire i punteggi osservati e i punteggi latenti e di conseguenza la forma della relazione tra questi due tipi di punteggi.



PCA

✓ `sklearn`

Libreria per l'analisi dei dati

✓ `sklearn.decomposition`

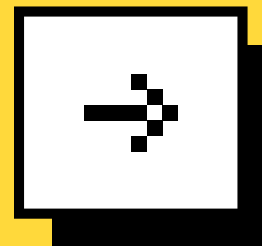
Modulo di algoritmi per la decomposizione di matrici

✓ `sklearn.decomposition.PCA`

Classe che prevede funzioni per calcolare le componenti principali e la trasformazione dei dati. Restituisce inoltre autovettori, autovalori e varianza spiegata.

MEAN SQUARED ERROR

`sklearn.metrics`



Ricostruzione dei punteggi osservati

$$MSE = \frac{\sum (O - \hat{O})^2}{n}$$



O indica i punteggi osservati



\hat{O} indica i punteggi ricostruiti



n è la numerosità campionaria



Ricostruzione della relazione tra
item e fattore (coordinate spaziali)






$$MSE = \frac{\sum (O - \hat{O})^2 + (\xi - PC)^2}{n}$$

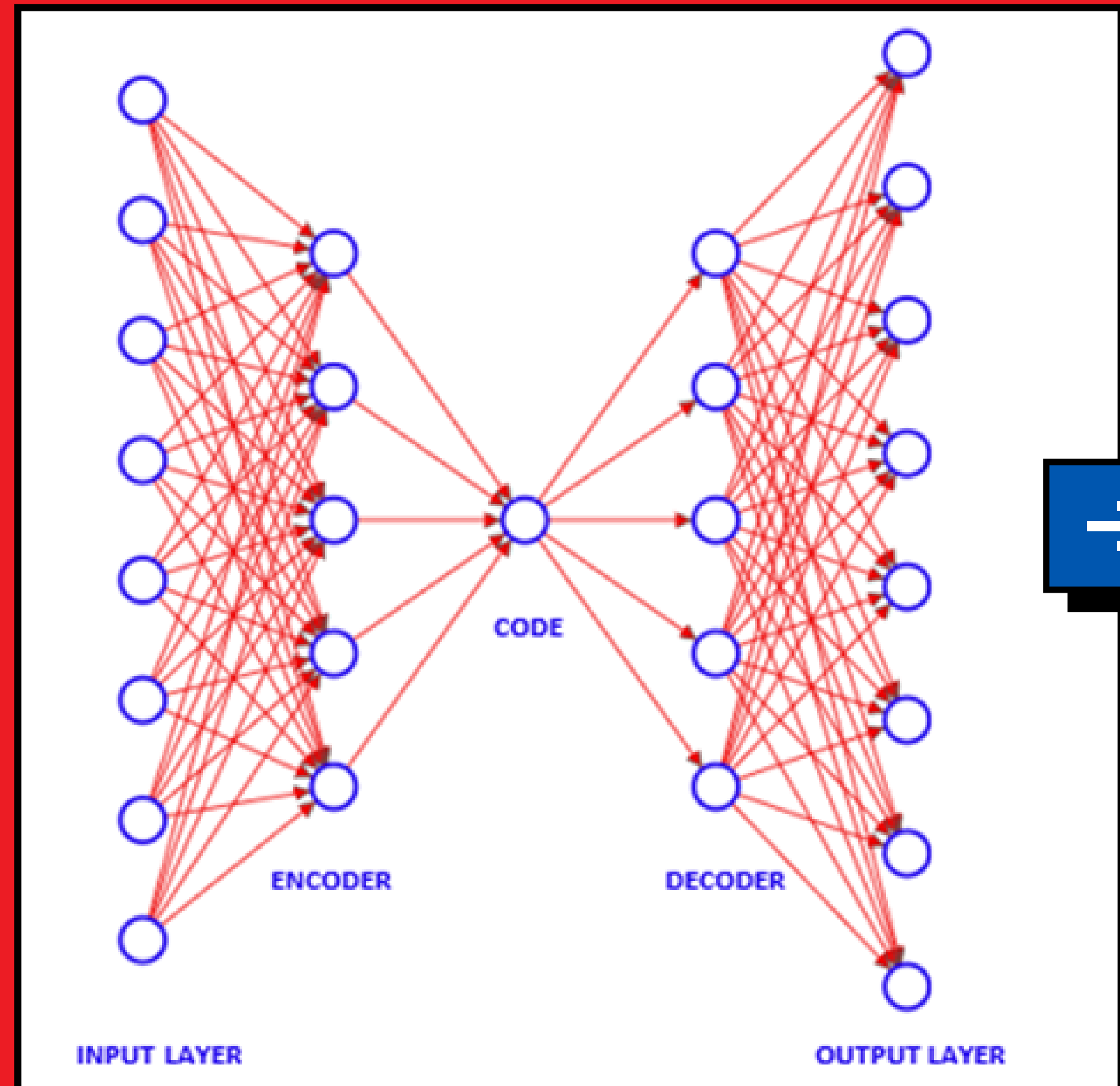


La componente principale è considerata
ricostruzione della variabile latente

Autoencoder

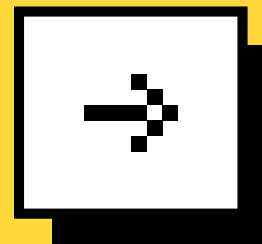
Architettura

-  Strato di input
-  Encoder a 5 nodi con attivazione sigmoidale
-  Strato interno a 1 nodo con attivazione lineare
-  Decoder a 5 nodi con attivazione sigmoidale
-  Strato di output con attivazione lineare



MEAN SQUARED ERROR

`sklearn.metrics`



Ricostruzione dei punteggi osservati

$$MSE = \frac{\sum (O - \hat{O})^2}{n}$$



O indica i punteggi osservati



\hat{O} indica i punteggi ricostruiti

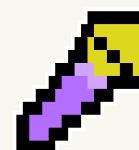


n è la numerosità campionaria



Ricostruzione della relazione tra
item e fattore (coordinate spaziali)

$$MSE = \frac{\sum (O - \hat{O})^2 + (\xi - CODE)^2}{n}$$



L'output dello strato interno è considerato
ricostruzione della variabile latente

Campioni

I dataset ($n=800$, $k=8$, $D=1$) sono stati simulati con Python,
seguendo le indicazioni di Bauer.

01

Generazione
delle variabili
latenti



02




Scelta della
funzione che lega le
variabili latenti ai
punteggi osservati

03

Generazione delle
variabili
osservate

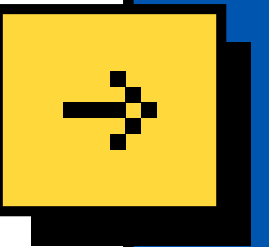



$$\xi = k + \sqrt{\varphi} \cdot s$$


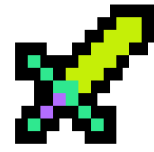
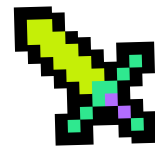
-  k è la media fattoriale ed è sempre uguale a 0;
-  φ è la varianza fattoriale ed è sempre uguale a 1;
-  s è una variabile casuale normale standardizzata.

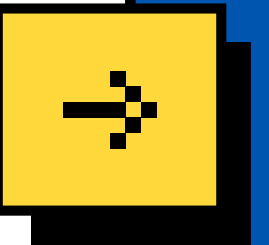
Punteggi fattoriali




Sono stati simulati in modo che seguissero una distribuzione normale standardizzata e che rientrassero tra -4 e +4.






$$y = t + \lambda \cdot \xi + \varepsilon$$




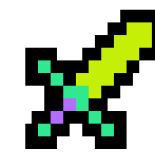

-  t è l'intercetta ed è uguale a 0 in tutti i casi;
-  λ è il coefficiente angolare ed è uguale a 0.8 in tutti i casi;
-  ε è l'errore di misurazione.

**Relazione lineare
tra item e fattore**

Modellata attraverso la funzione classica utilizzata nell'analisi fattoriale:
la funzione di una retta con l'aggiunta dell'errore di misurazione.





Errore di misura


$$\varepsilon = \sqrt{\psi} \cdot s$$




ψ è la varianza dei residui

Varianza dei
residui




$$\psi = \frac{(1 - r_{tt}) \cdot \varphi}{r_{tt}}$$


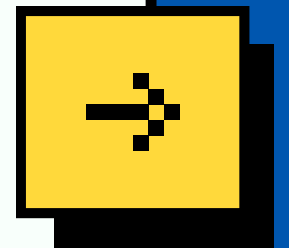






r_{tt} è l'attendibilità del test ed è posta a 0.8






$$z = 0.3 \cdot \left(\left(\frac{1}{(1 + e^{-a(\xi - b)}) \cdot 2} - 1 \right) \cdot 10 + \lambda \cdot \xi + \varepsilon \right)$$




-  e è il numero di Nepero;
-  a è la pendenza della sigmoide;
-  b è il punto di flesso della sigmoide;
-  i moltiplicatori 0.3, 2 e 10 fanno rientrare i punteggi tra -4 e +4.

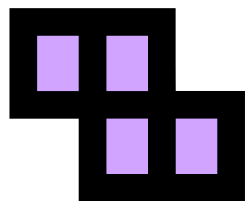
Relazione non lineare
tra item e fattore

Modellata attraverso la funzione di una sigmoide, con l'aggiunta di un elemento lineare.

Differenze tra i dataset

I dataset si differenziavano tra loro per le seguenti caratteristiche degli item:

01



Proporzione di item lineari e non lineari

- La proporzione di item lineari poteva variare tra 0/8 e 7/8
- La proporzione di item non lineari poteva variare tra 1/8 e 8/8



02

Parametri degli item non lineari

- Uguali per tutti gli item del dataset
- Uguali a gruppi di item

$a = [0.5, 1, 2, 3];$

$b = [-3, -2, -1, 1, 2, 3].$

Dataset 1

01

0 ITEM LINEARI

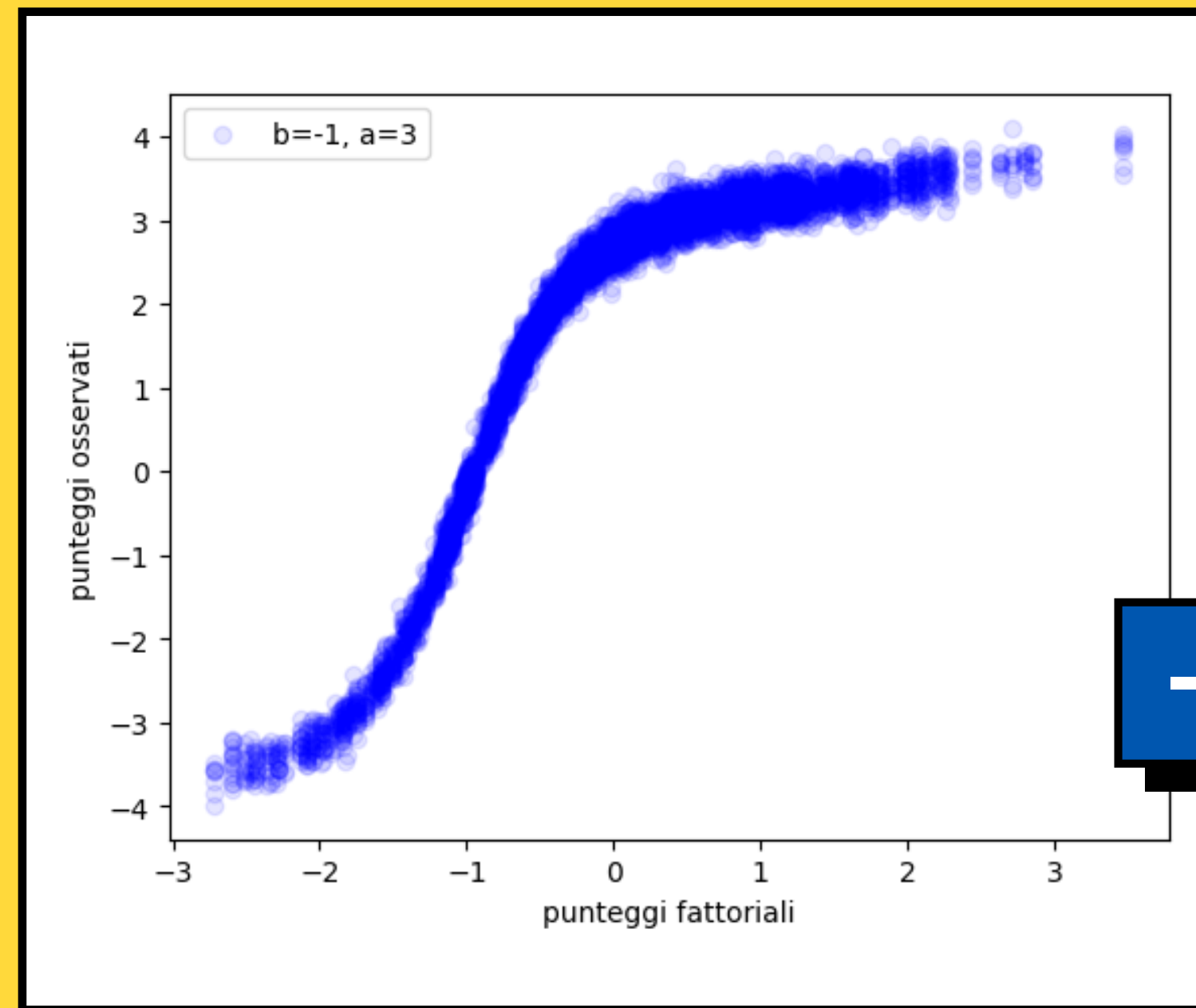
8 ITEM NON LINEARI

02

Parametri uguali per tutti
gli item

$a = 3$

$b = -1$



Dataset 2

01

7 ITEM LINEARI

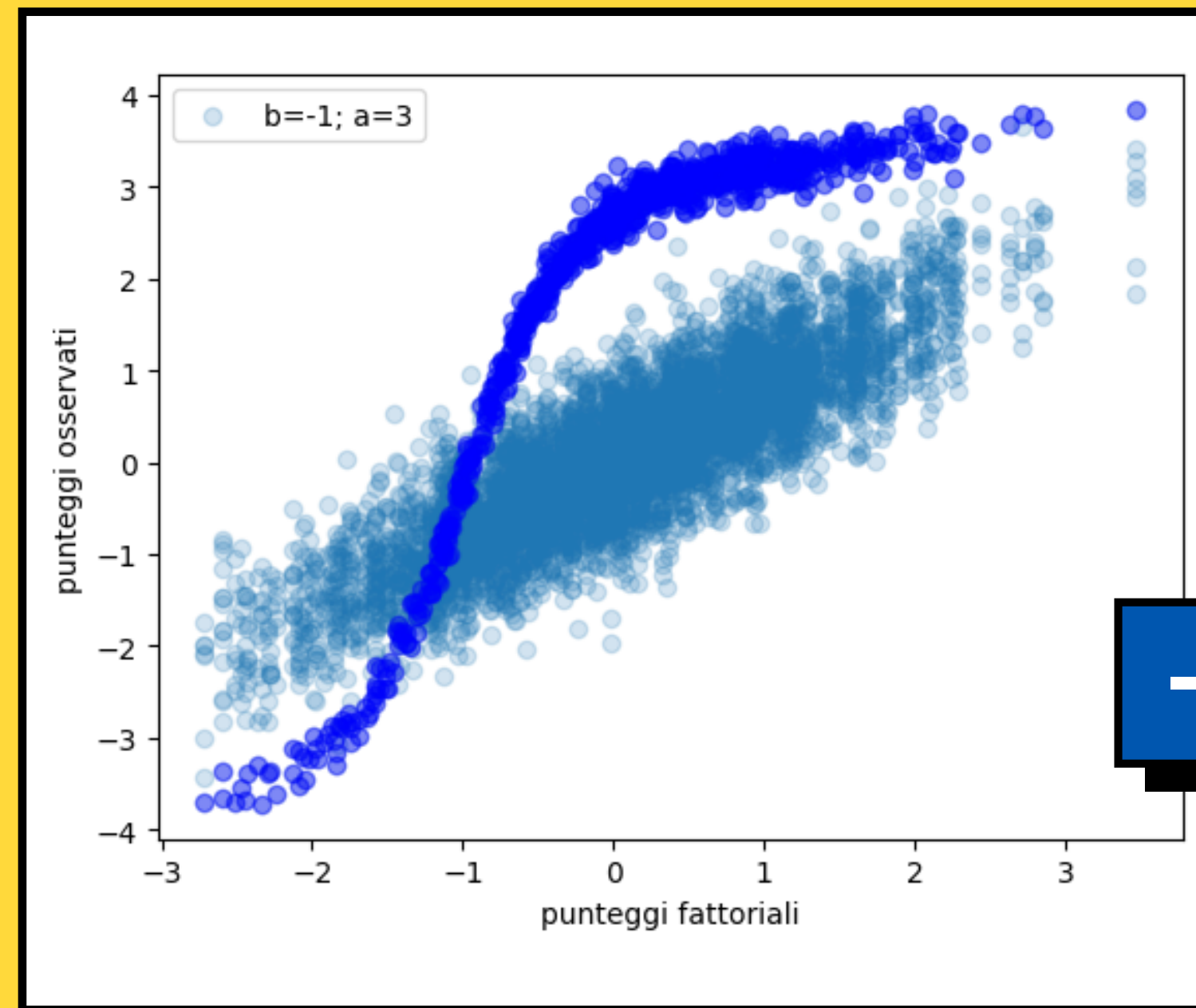
1 ITEM NON LINEARE

02

Parametri dell'item non
lineare

$$a = 3$$

$$b = -1$$



Dataset 3

01

0 ITEM LINEARI

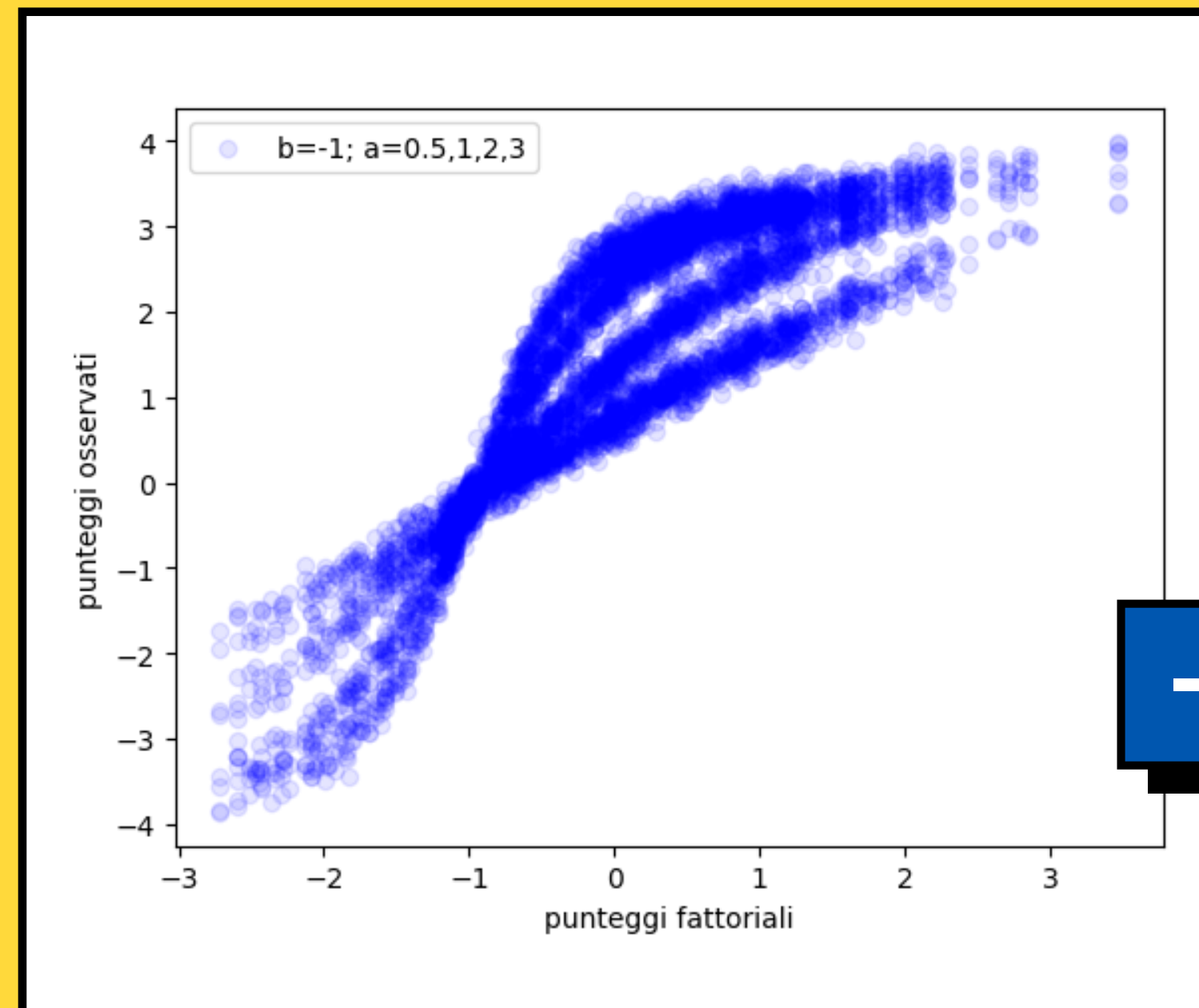
8 ITEM NON LINEARI

02

Pendenze degli item non
lineari uguali a 2 a 2

$a = [0.5, 1, 2, 3]$

$b = -1$



Dataset 4

01

0 ITEM LINEARI

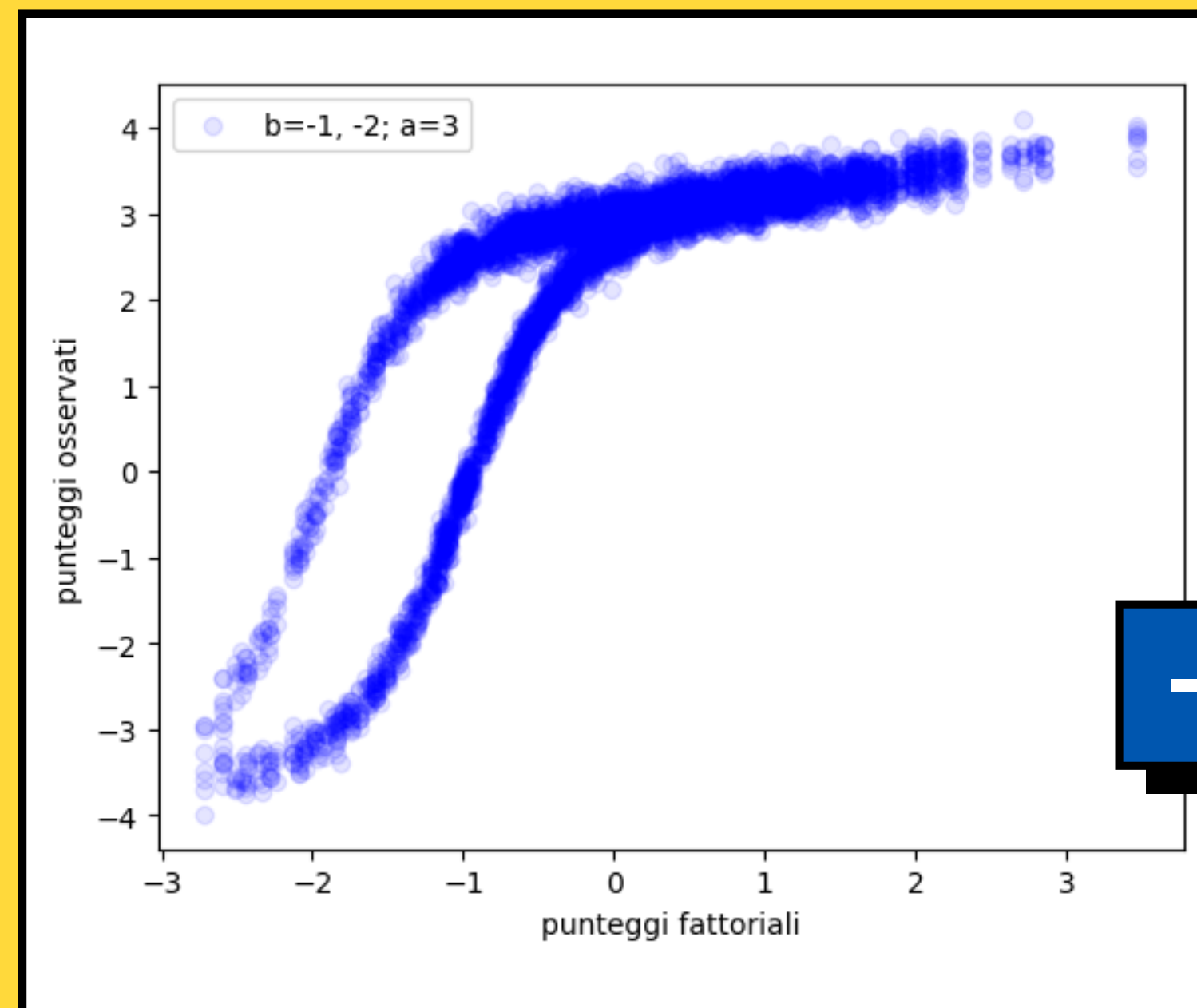
8 ITEM NON LINEARI

02

Punti di flesso degli item
non lineari uguali a 4 a 4

$a = 3$

$b = [-2, -1]$



Dataset 5

01

0 ITEM LINEARI

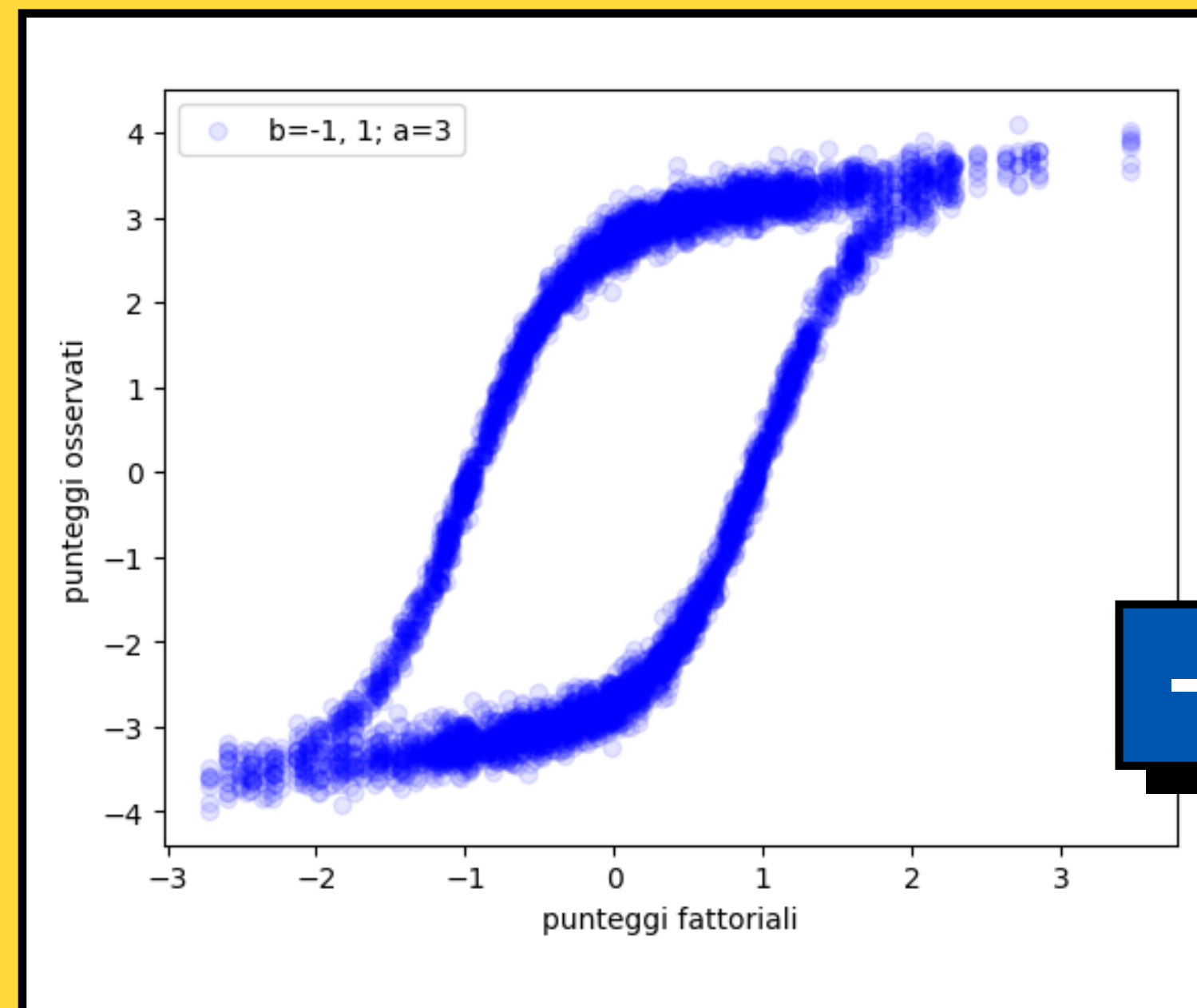
8 ITEM NON LINEARI

02

Punti di flesso degli item
non lineari uguali a 4 a 4

$a = 3$

$b = [-1, 1]$




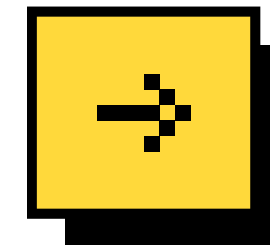
Risultati

 LIVELLI 

★ Totalità degli item

★ Singolo item non lineare

 Errori di
ricostruzione dei
punteggi osservati





Dataset 1

MSE punteggi osservati

Totalità degli item

Autoencoder

0.02

PCA

0.02

MSE punteggi osservati

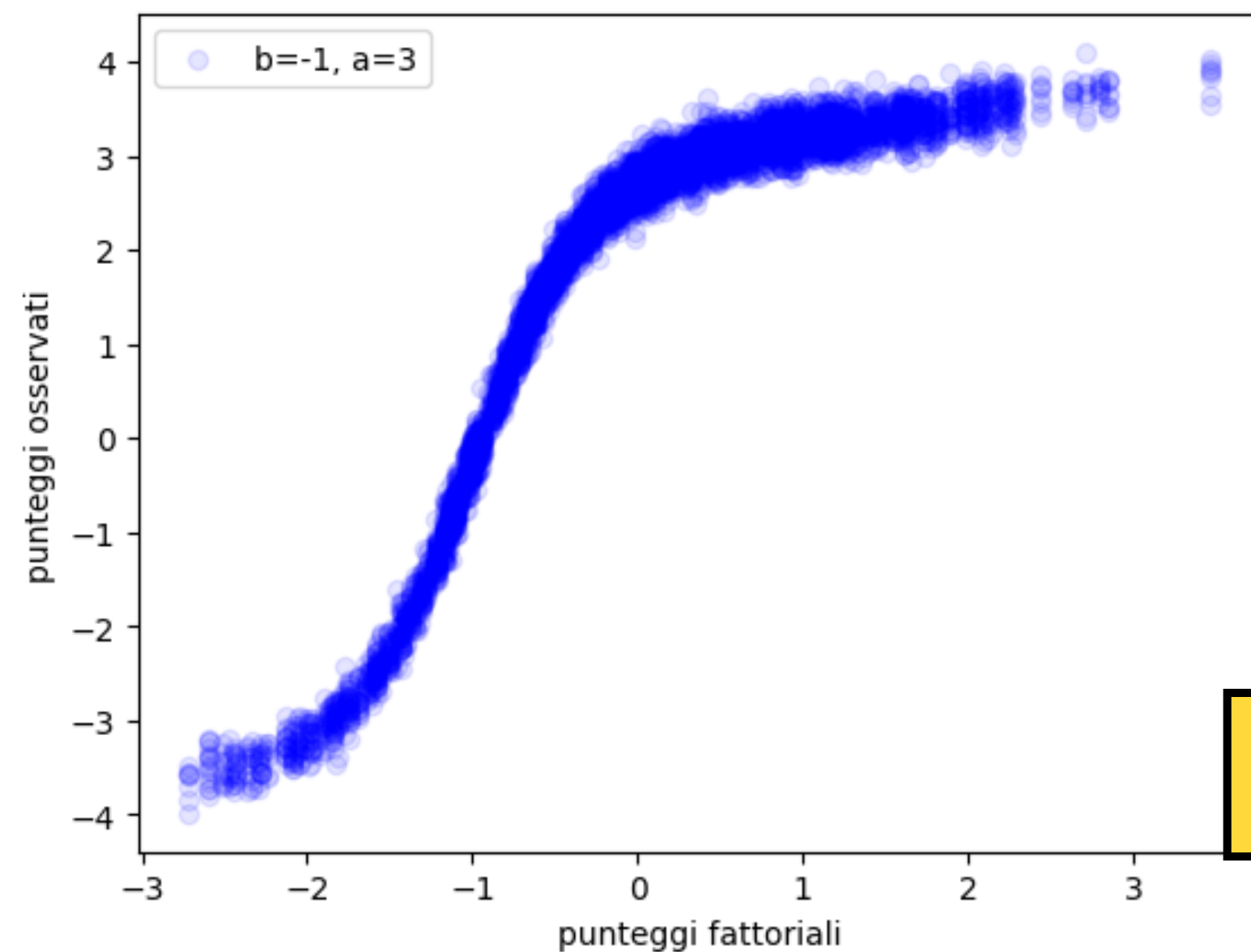
Singolo item non lineare

Autoencoder

0.05

PCA

0.06





Dataset 2

MSE punteggi osservati

Totalità degli item

Autoencoder

0.2

PCA

0.2

MSE punteggi osservati

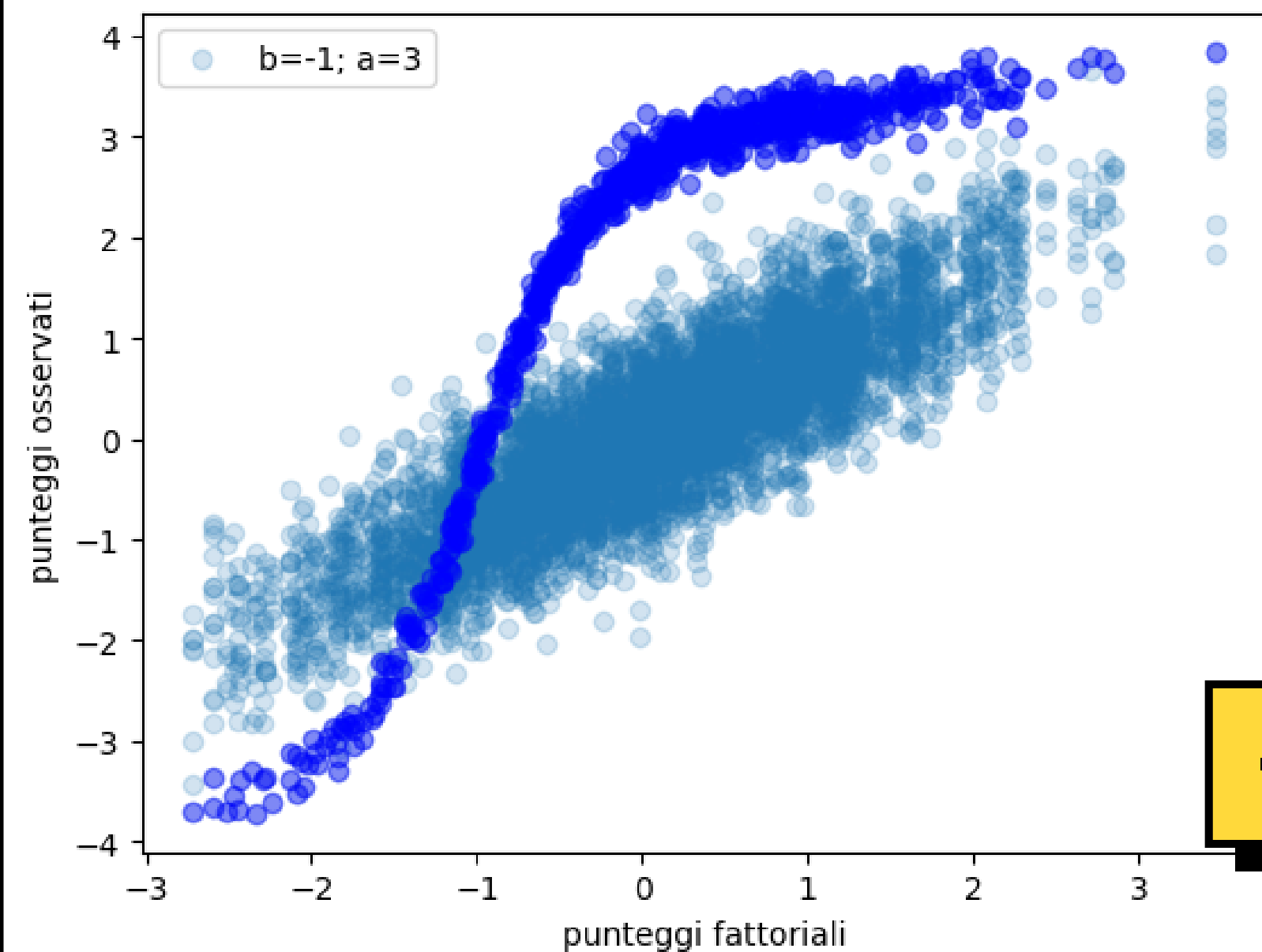
Singolo item non lineare

Autoencoder

0.04

PCA

0.36





Dataset 3

MSE punteggi osservati

Totalità degli item

Autoencoder

0.02

PCA

0.07

MSE punteggi osservati

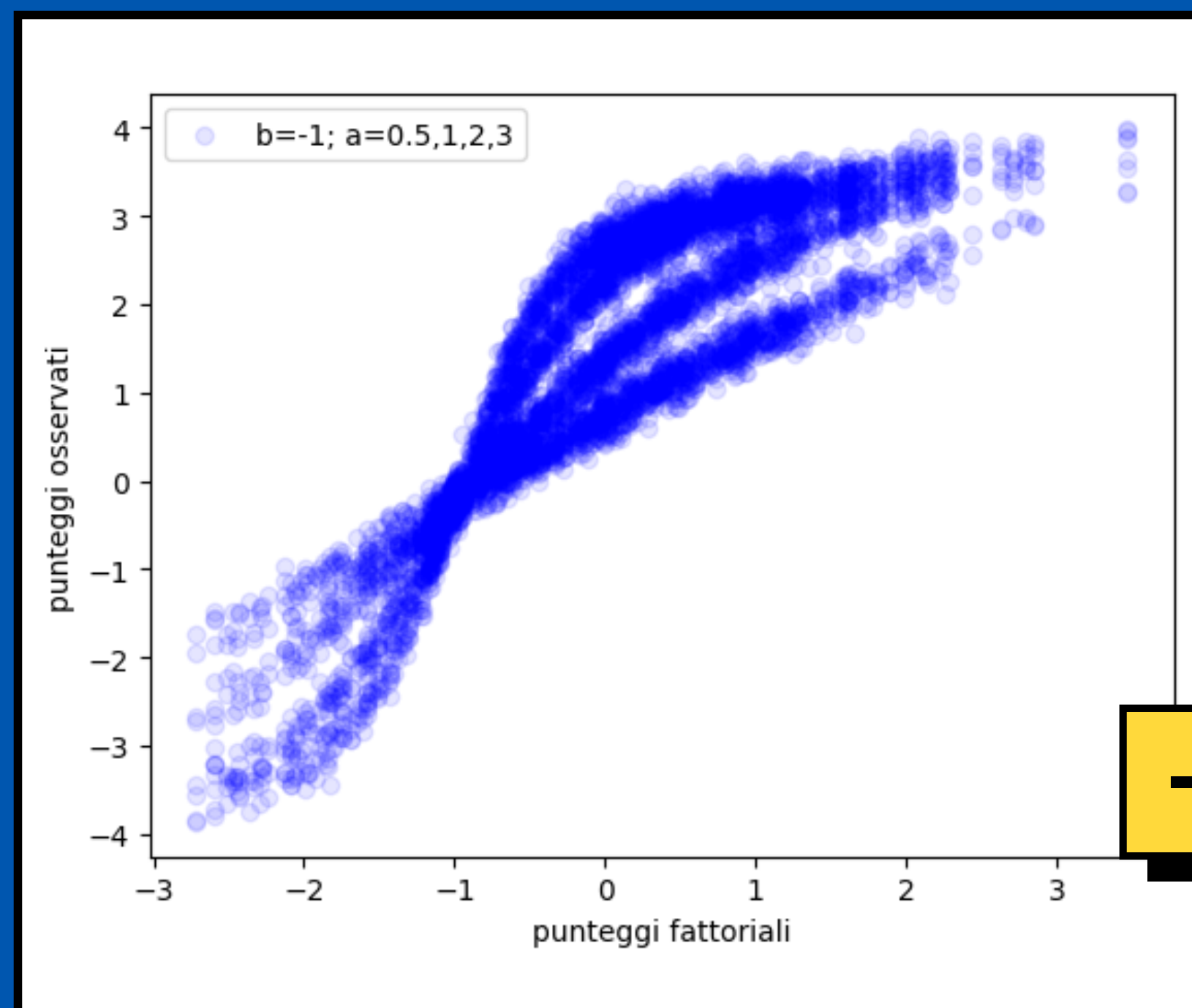
Singolo item non lineare

Autoencoder

0.03

PCA

0.09





Dataset 4

MSE punteggi osservati

Totalità degli item

Autoencoder

0.02

PCA

0.1

MSE punteggi osservati

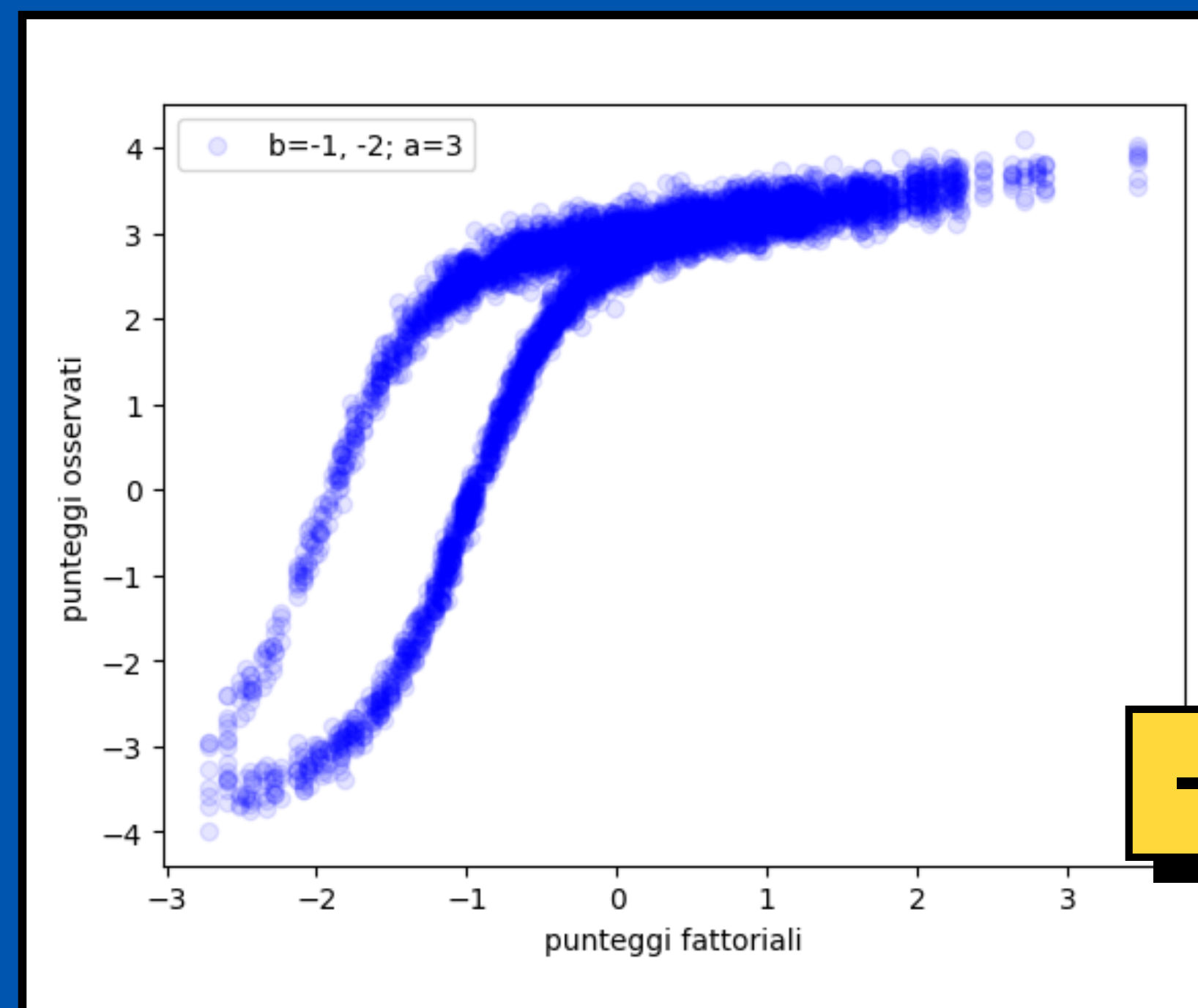
Singolo item non lineare

Autoencoder

0.03

PCA

0.2





Dataset 5

MSE punteggi osservati

Totalità degli item

Autoencoder

0.02

PCA

0.7

MSE punteggi osservati

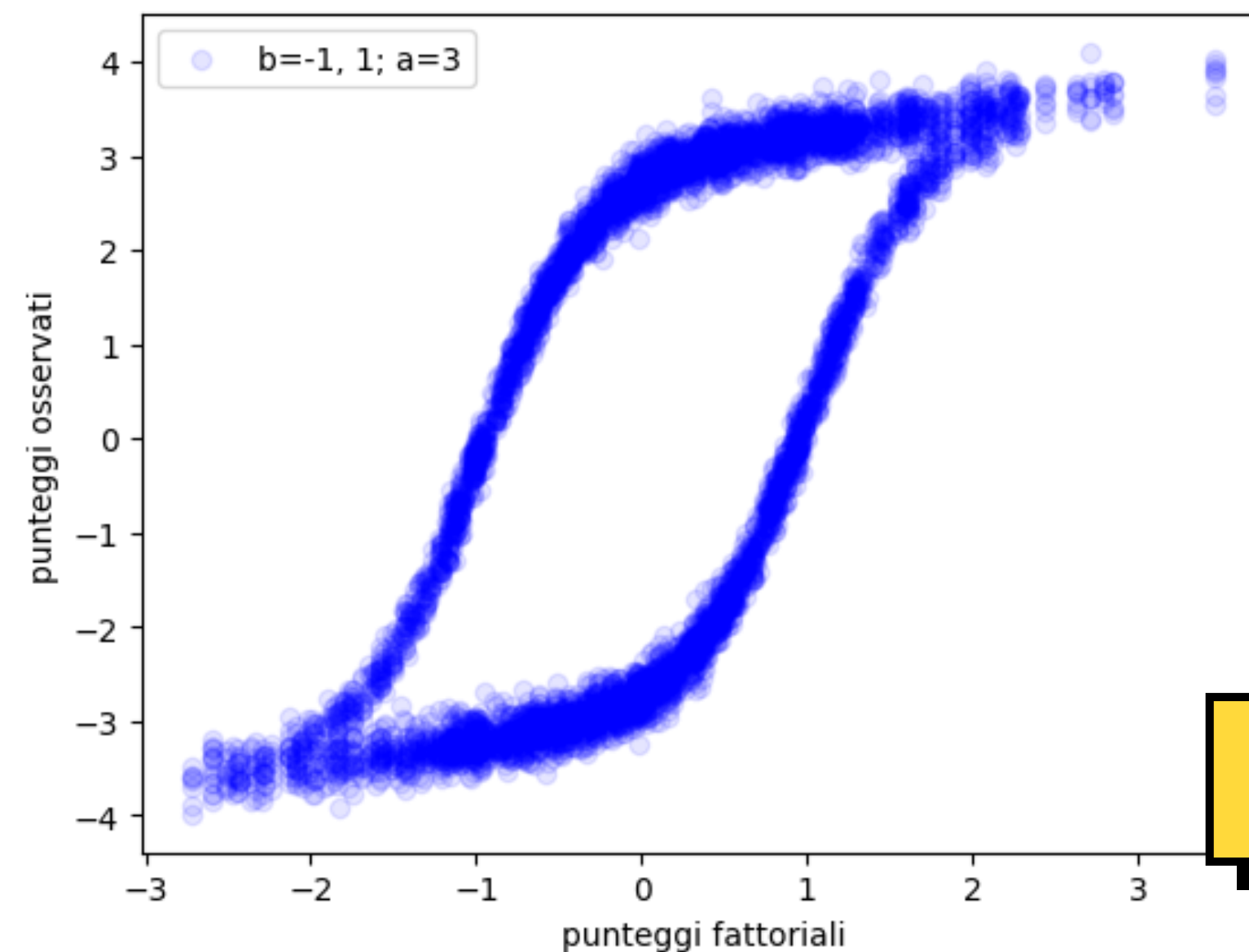
Singolo item non lineare

Autoencoder

0.02


PCA


0.7

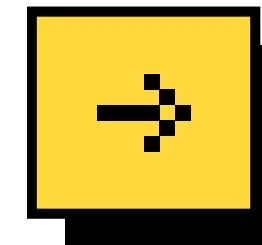


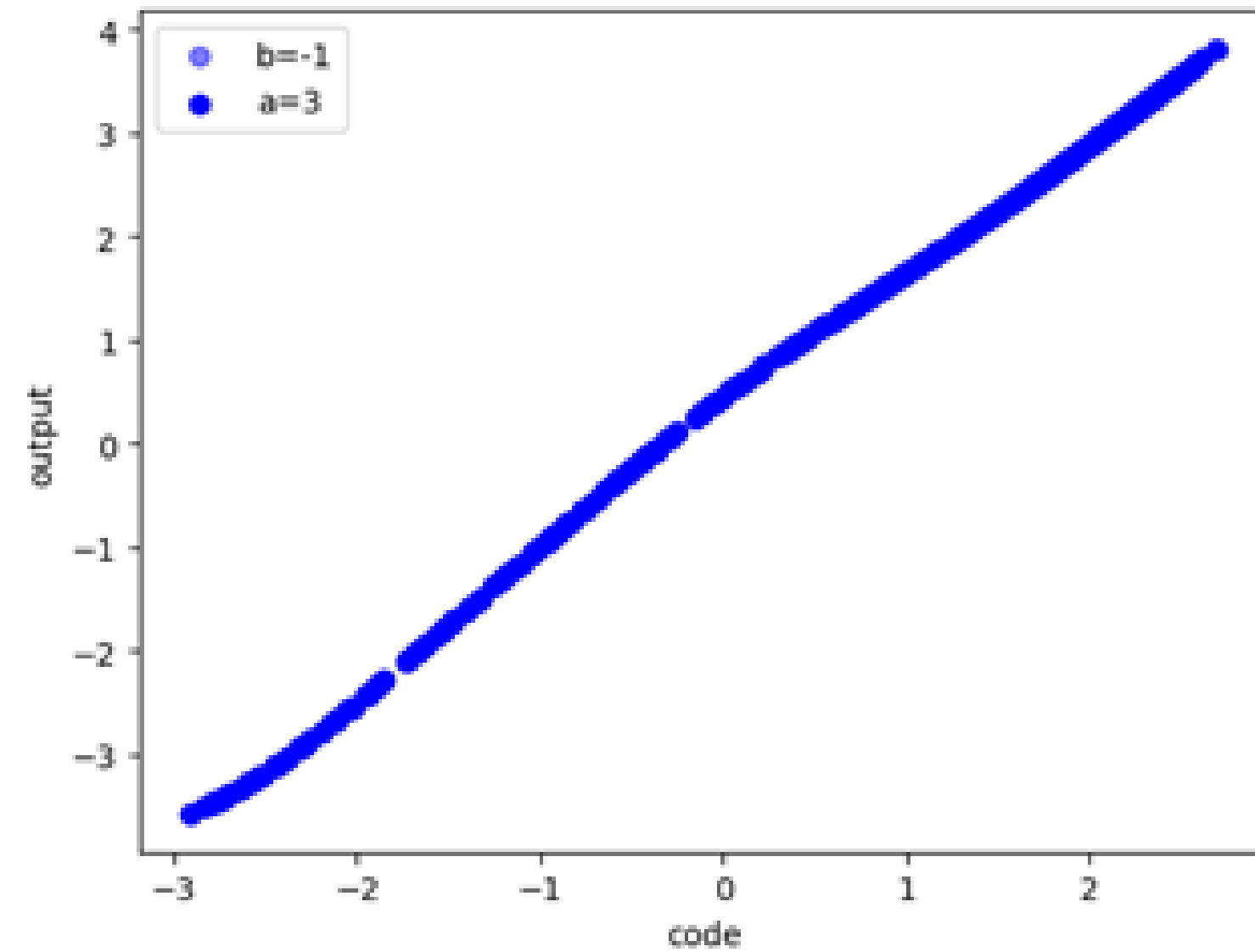
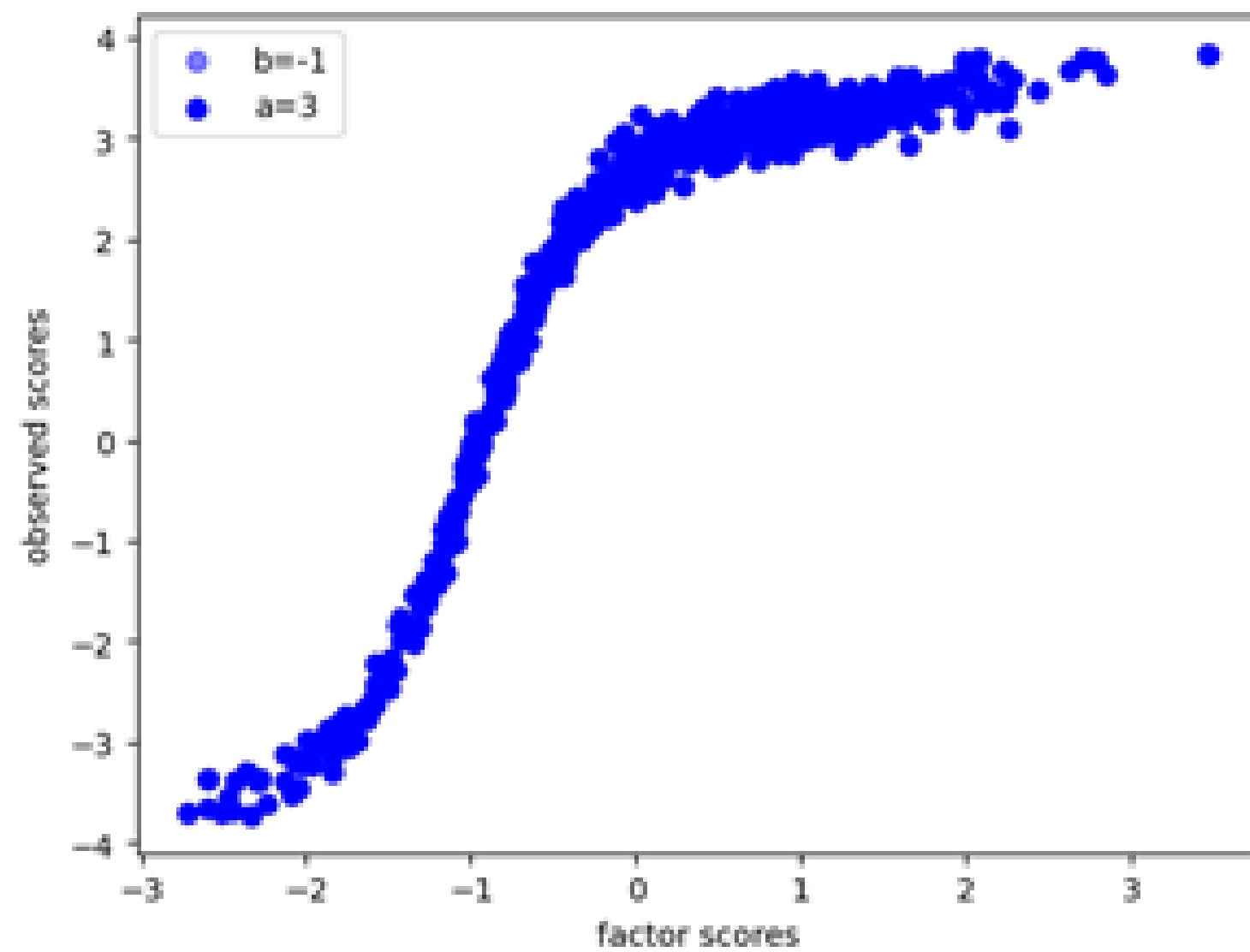
Risultati

 LIVELLI 

 Singolo item non
lineare

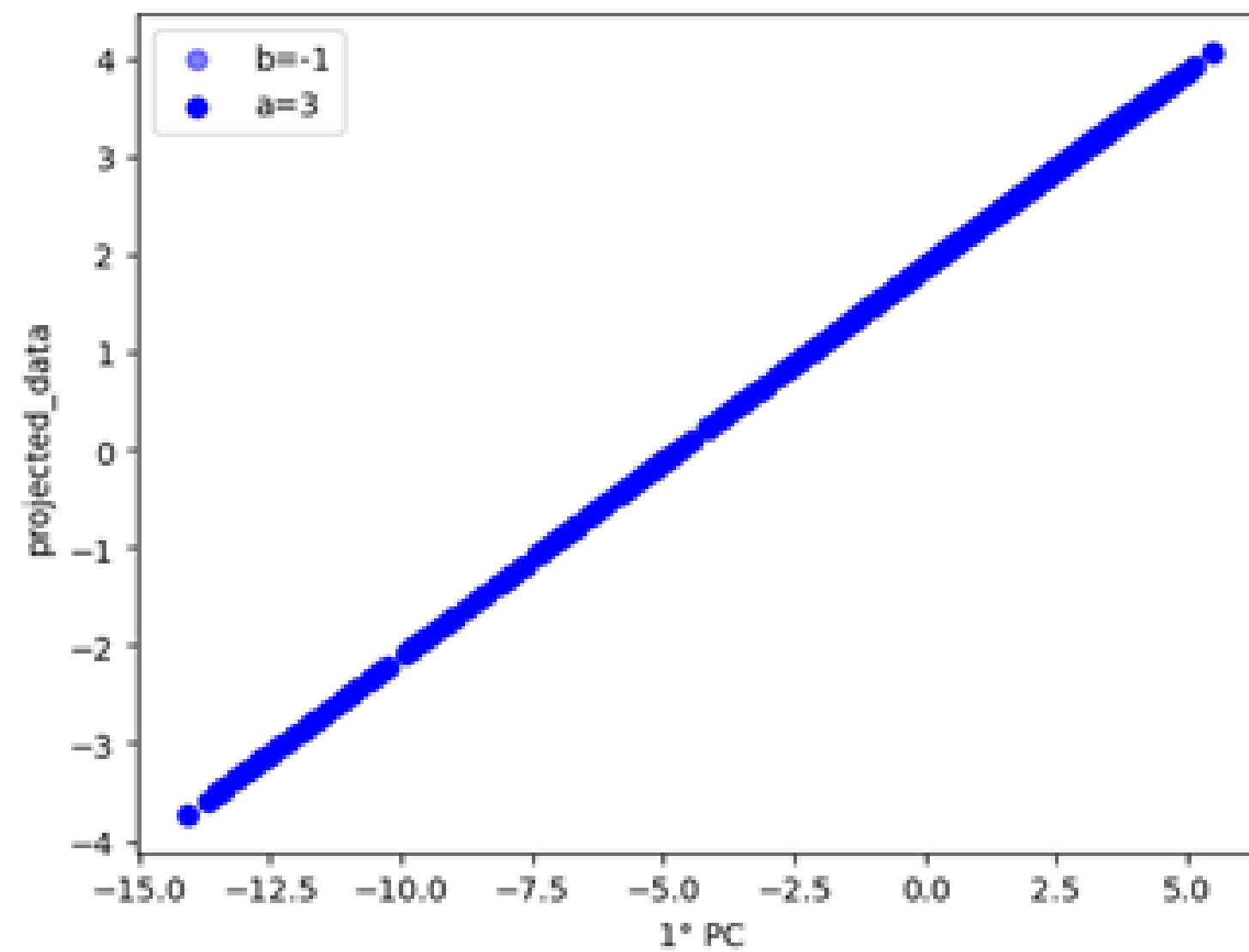
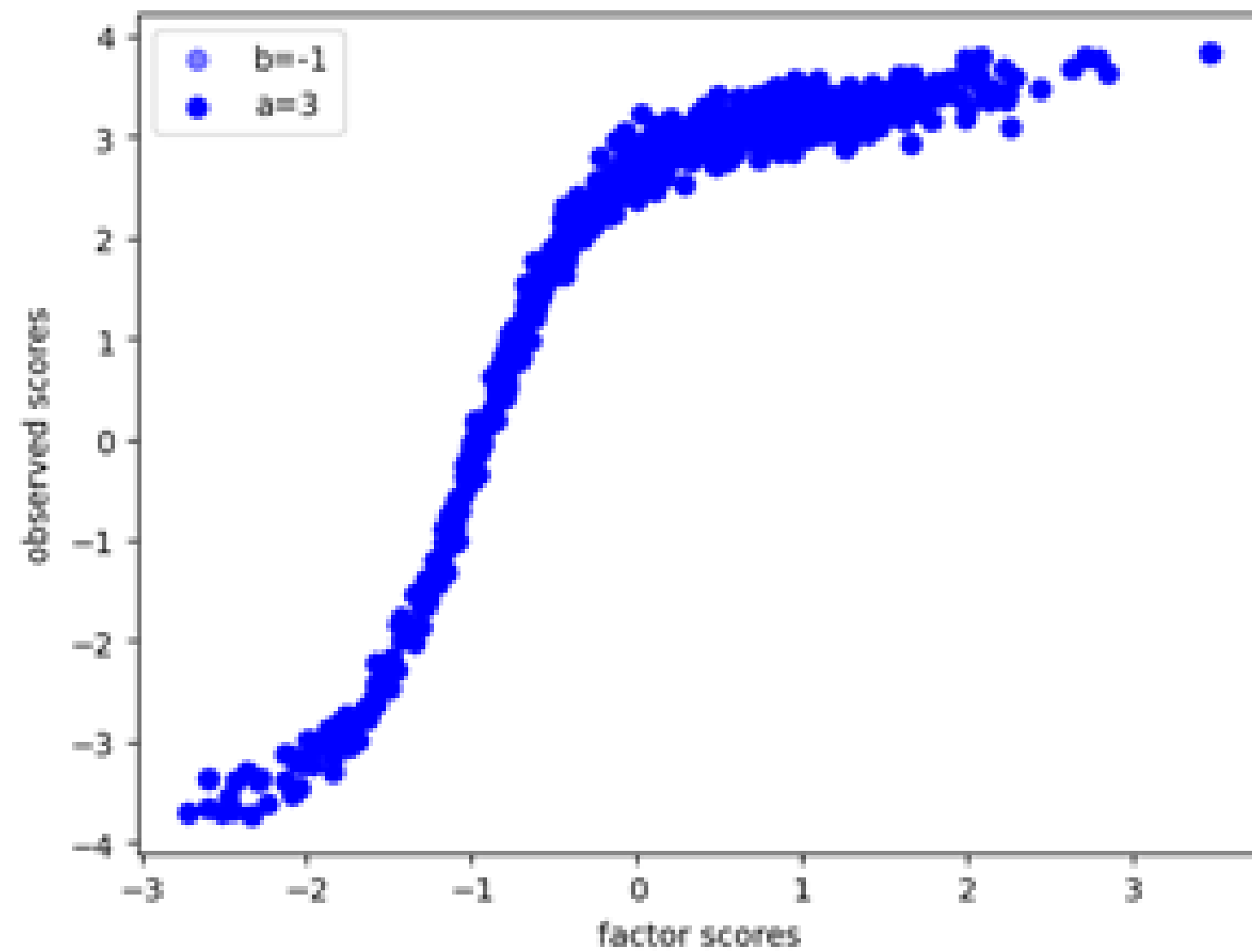
 Errori di
ricostruzione della
relazione tra item e
fattore






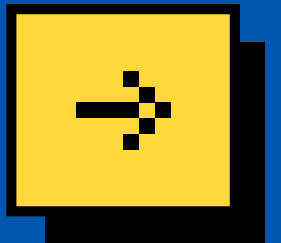
MSE coordinate spaziali
Autoencoder: 0.4

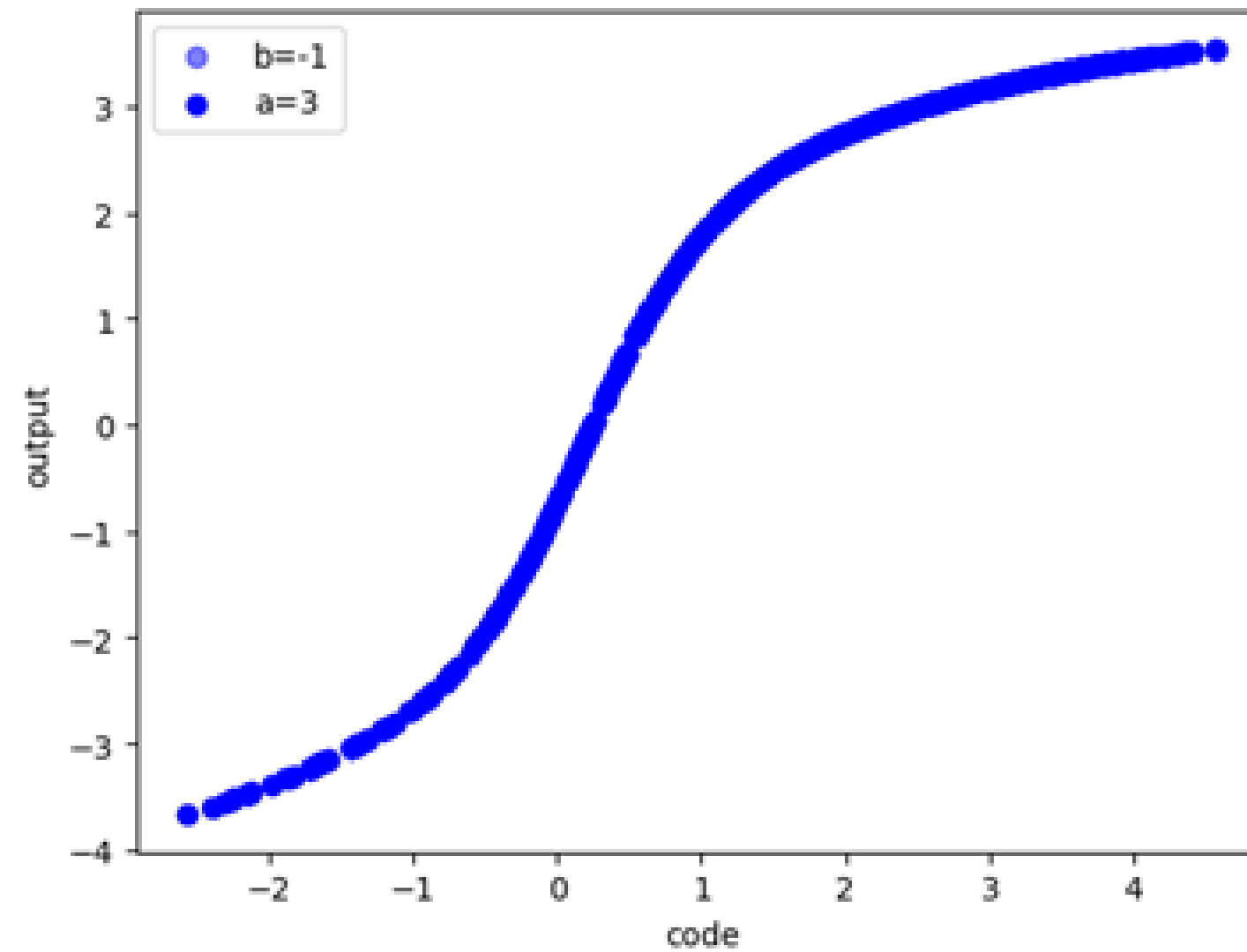
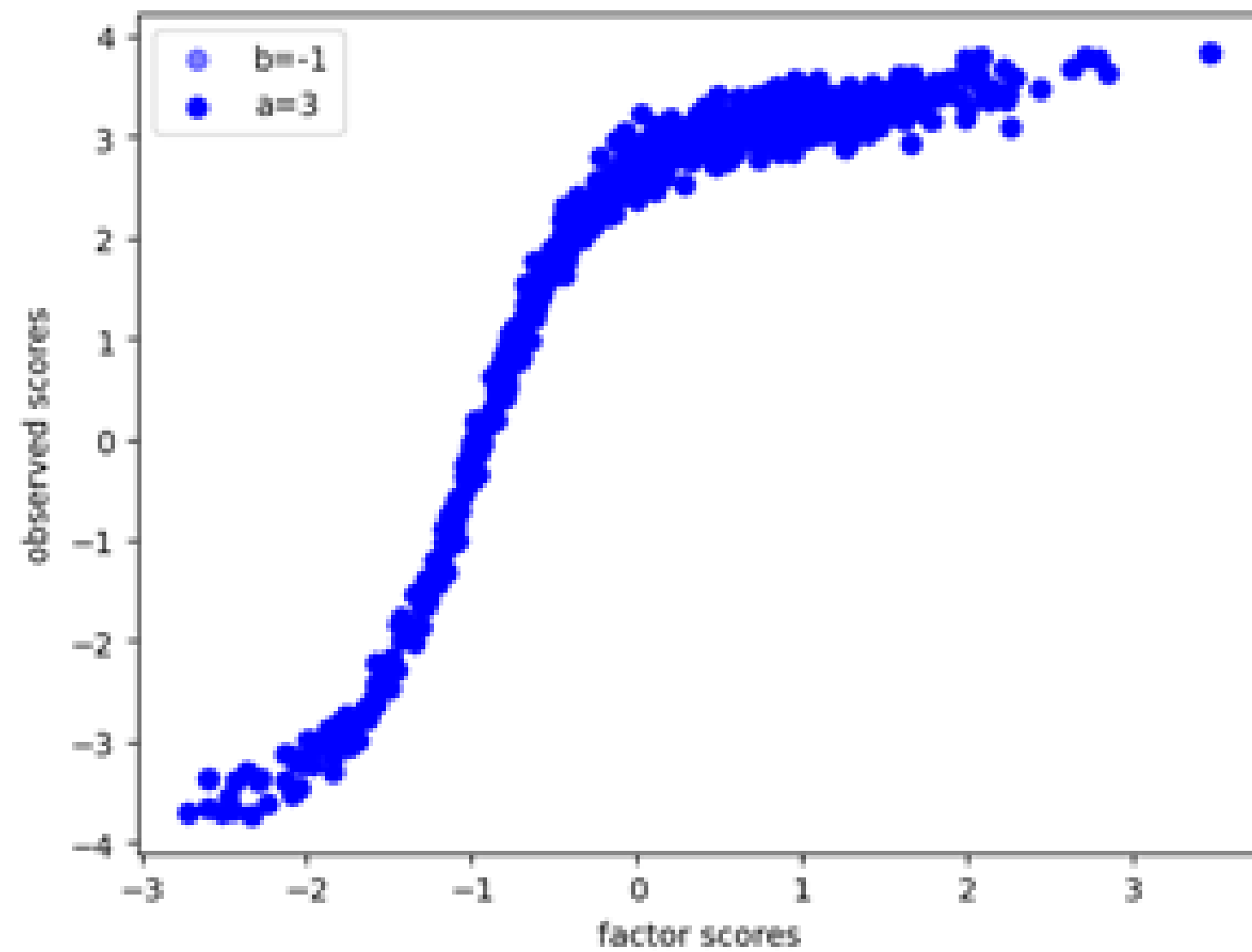
Dataset 1



 MSE coordinate spaziali 
PCA: 7.4

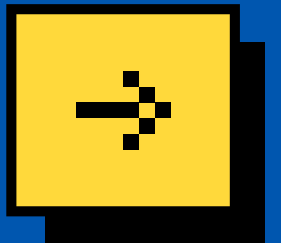
Dataset 1

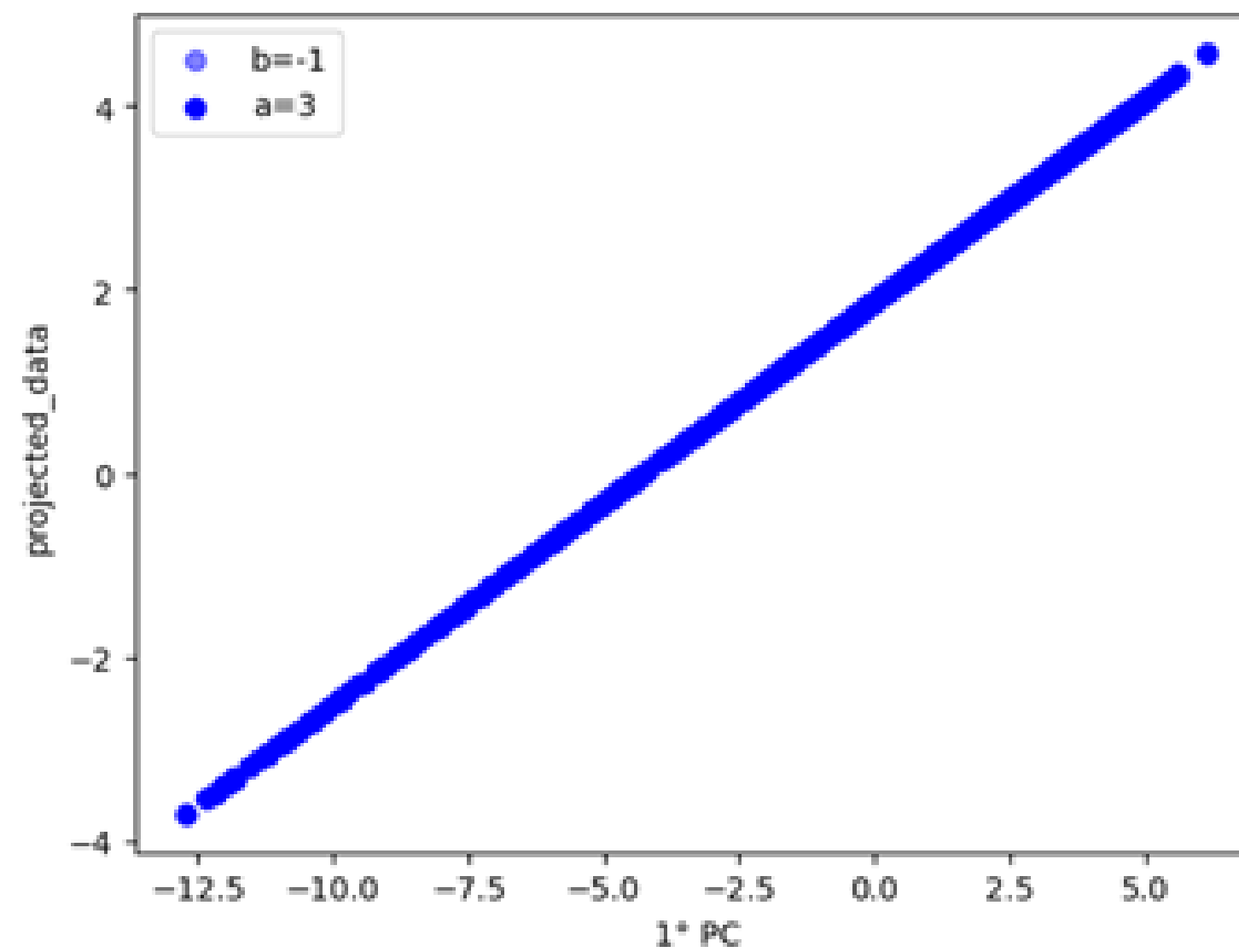
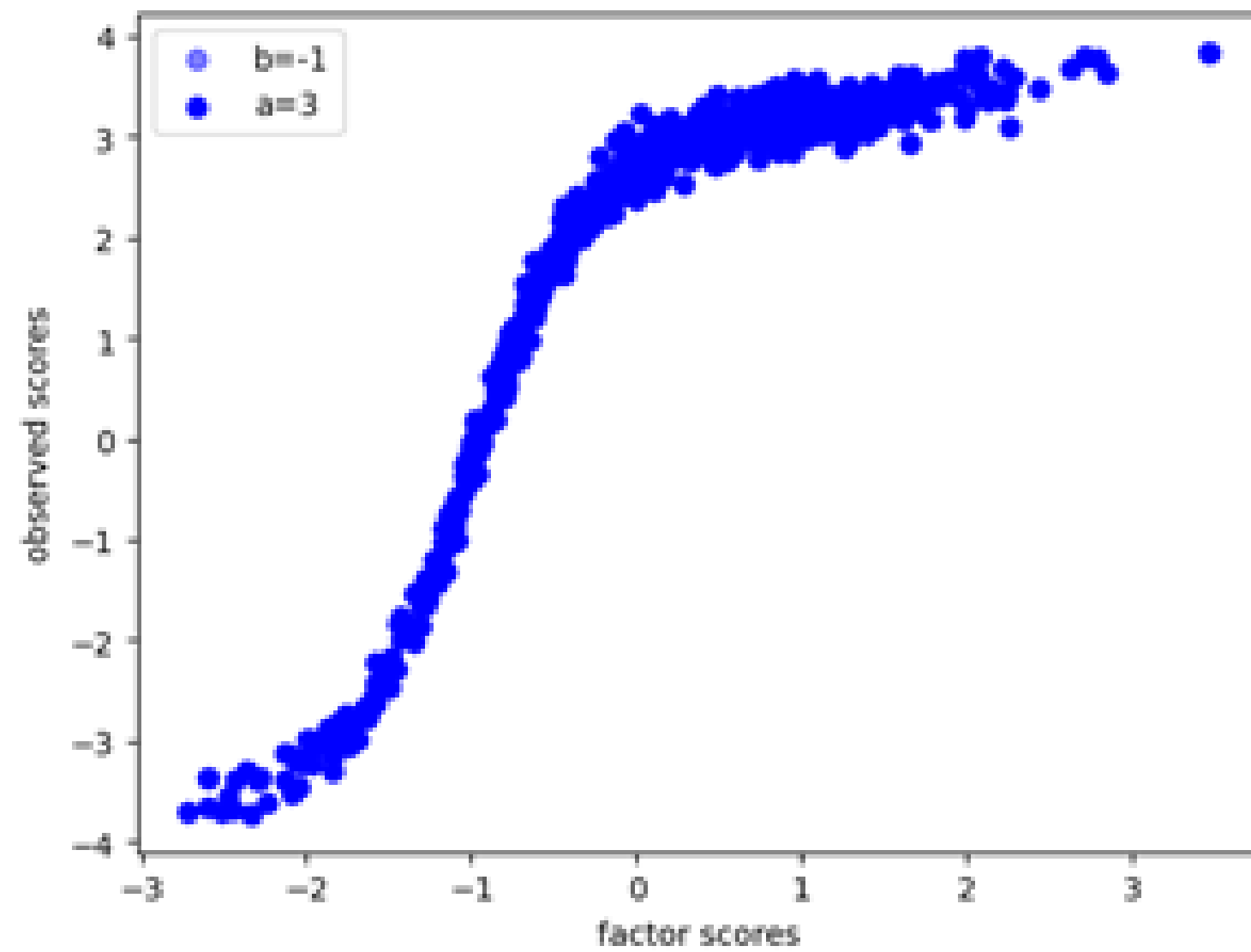






⌘ MSE coordinate spaziali ⌘
Autoencoder: 0.6

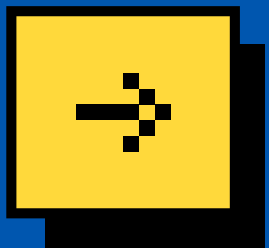
Dataset 3

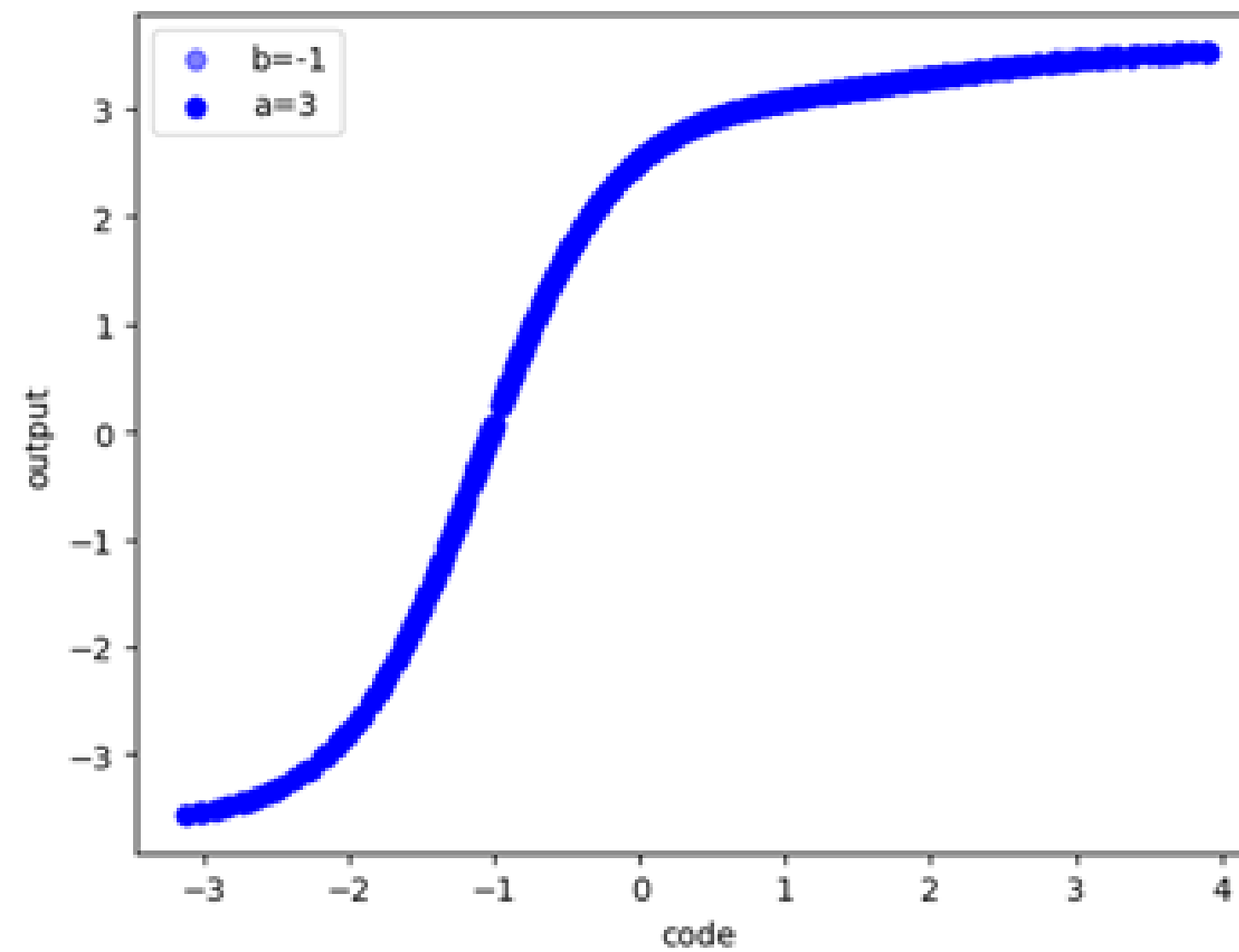
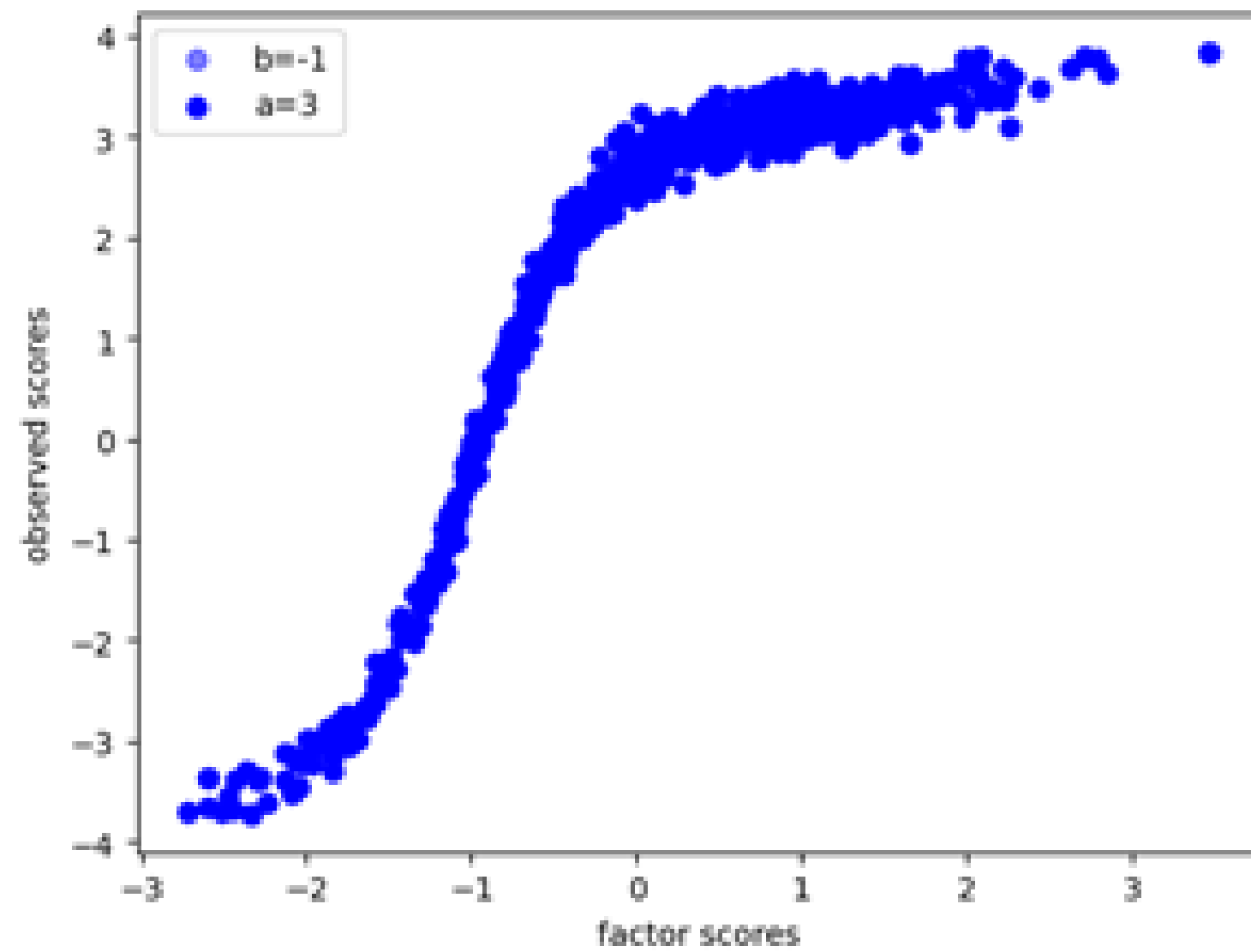




 MSE coordinate spaziali 
PCA: 5.5

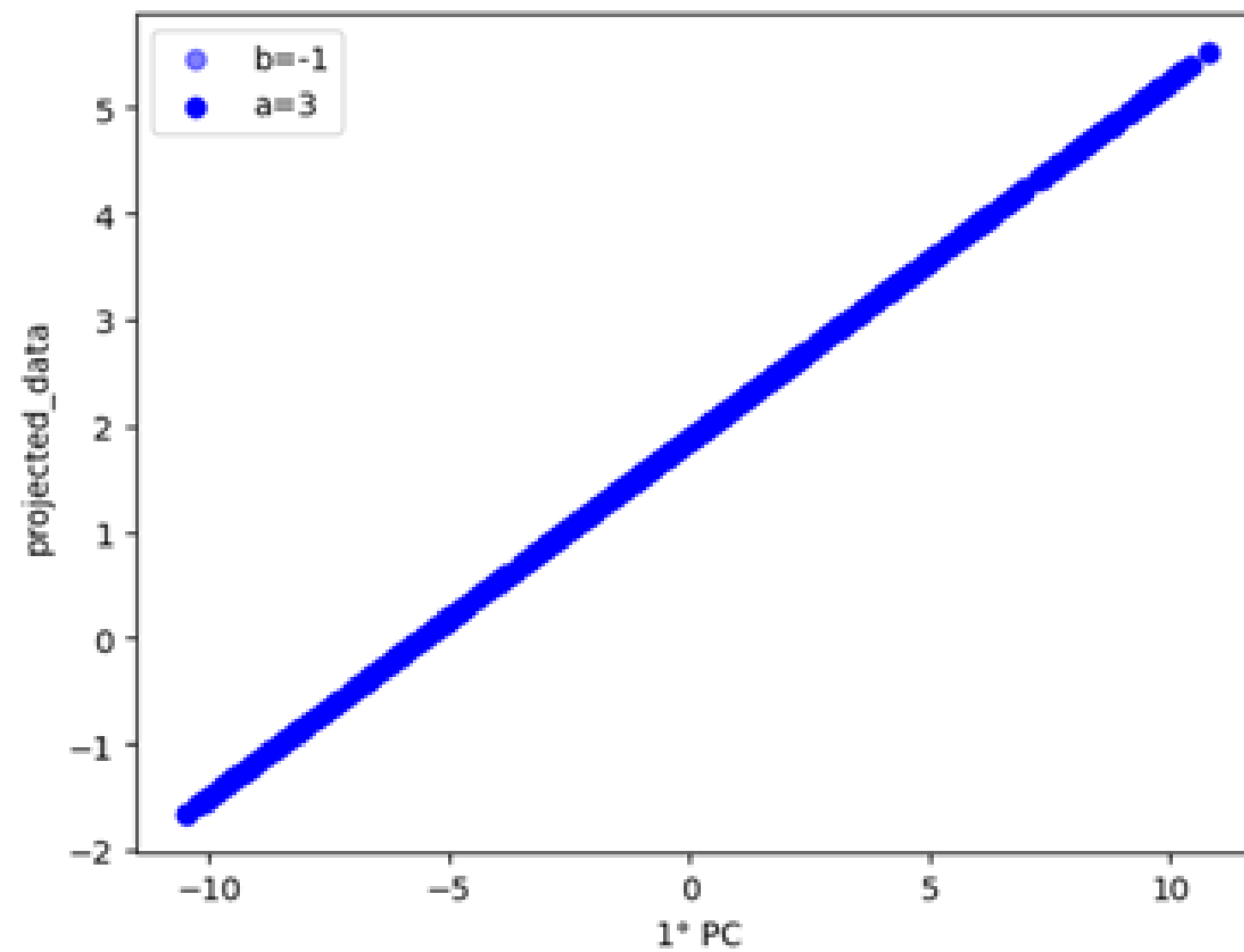
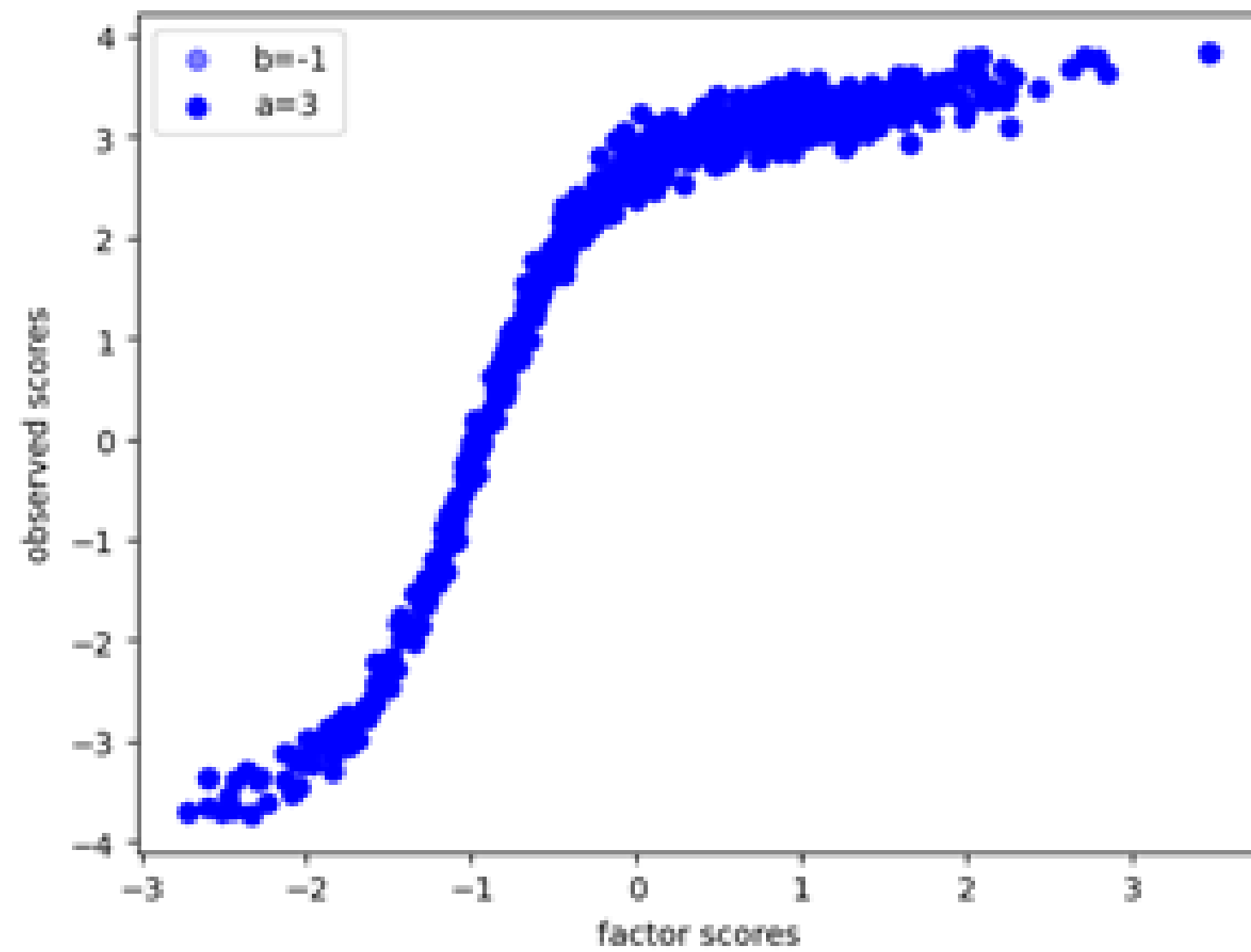
Dataset 3





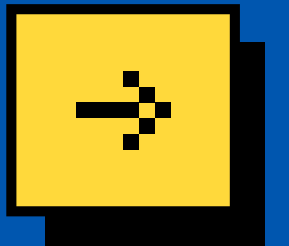
⌘ MSE coordinate spaziali ⌘
Autoencoder: 0.2

Dataset 5

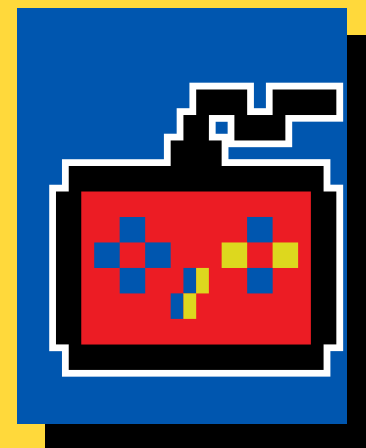
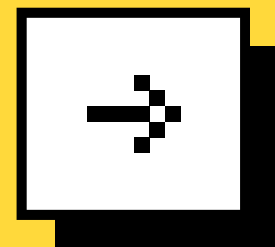


 MSE coordinate spaziali 
PCA: 6.8

Dataset 5




Oltre la
ricostruzione dei
punteggi
osservati



L'autoencoder riproduce uno spazio che si avvicina molto di più a quello simulato rispetto a quello riprodotto dalla PCA, non appena c'è un elemento di differenziazione tra gli item nel dataset.

Per riuscire in un tale compito, è indispensabile ricostruire accuratamente non solo le ordinate del grafico della funzione, ma anche le ascisse ossia i valori associati alla variabile latente.

La capacità dei metodi di riduzione della dimensionalità lineari di ridurre la dimensionalità delle osservazioni senza perdere troppa informazione su di esse non esclude una perdita importante:

 L'informazione sul **tipo di interazione** tra le **dimensioni dello spazio**, che rappresentano le variabili psicologiche, e **gli elementi di questo spazio**, che rappresentano i comportamenti attuati dai soggetti sulla base di queste variabili.



Ciò significherebbe ottenere false conoscenze sulle **quantità di tratto** possedute dai soggetti e identificare in modo errato i **comportamenti da aspettarsi dagli individui** sulla base del loro tratto latente.



♥ GRAZIE ♥



RIFERIMENTI

- ♥ Bauer, D. J. (2005). The Role of Nonlinear Factor-to-Indicator Relationships in Tests of Measurement Equivalence. *Psychological Methods*, 10(3), 305-316.
- ♥ Belzak, W. C. M., & Bauer, D. J. (2019). Interaction effects may actually be nonlinear effects in disguise: A review of the problem and potential solutions. *Addictive Behaviors*, 94, 99-108.