

# I: PREZENTARE GENERALĂ

- **Titlu:** „Învățare supervizată – regresie”
- **Obiective:**
  - Conceptul de regresie
  - Modele de regresie (Liniară, Polinomială și Arbori)
  - Supraadaptare (overfitting) și tehnici de regularizare
  - Metricile de evaluare ( $R^2$ , MSE, MAE)
  - Exemple de utilizare în lumea reală

# 2: INTRODUCERE ÎN REGRESIE

- **Ce este regresia?**

- Prezicerea unei valori numerice continue pe baza variabilelor de intrare
- În contrast cu clasificarea (unde ieșirile sunt discrete)

- **Utilizări în lumea reală:**

- Prognoza vânzărilor
- Prezicerea prețurilor locuințelor
- Estimarea costurilor de producție

- **Exemplu:** Prezentați un scenariu scurt despre prezicerea prețurilor locuințelor pe baza suprafeței, locației etc.

# 3: BAZELE REGRESIEI LINIARE

- **Concept:**

- Forma modelului:  $(y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)$
- Presupune o relație liniară între predictorii și țintă

- **Terminologie:**

- Coeficienți (ponderi)
- Ordinată la origine (intercept)
- Reziduuri (erori)

- **Metoda pătratelor minime (OLS):**

- Minimizarea sumei pătratelor reziduurilor

- **Exemplu:** Set de date simplu cu o singură caracteristică (ex. ore de studiu) și o țintă (scor la examen), ilustrând linia de regresie

# 4: REGRESIA POLINOMIALĂ

- **Când liniarea nu este suficientă:**

- Relația dintre variabile poate fi neliniară

- **Forma modelului:**

- Introduce termeni polinomiali (ex.  $x^2$ ,  $x^3$ ) pentru o potrivire mai bună

- **Atenție:**

- Gradele polinomiale superioare pot duce la supraadaptare

- **Exemplu:** Diagramă cu puncte care formează o relație curbată (ex. viteza mașinii vs. distanța de frânare), demonstrând o potrivire polinomială

# 5: ARBORI DE REGRESIE

- **Conceptul arborilor de decizie:**

- Împarte datele în funcție de praguri pe caracteristici
- Produce predicții constant-pas cu pas (piecewise-constant)

- **Avantaje:**

- Ușor de interpretat
- Manevreează bine neliniaritatea și interacțiunile

- **Dezavantaje:**

- Se pot supraadapta dacă se cresc prea adânc

- **Exemplu:** Pașii unui arbore de regresie pentru prezicerea prețului unei locuințe: mai întâi împarte după locație, apoi după numărul de camere, apoi suprafața etc.

- **Articol**

# 6: SUPRAADAPTARE & REGULARIZARE

- **Definiția supraadaptării (Overfitting):**

- Modelul are performanțe bune pe datele de antrenament, dar slabe pe date nevăzute

- **Cauze comune:**

- Prea multe caracteristici
- Regularizare insuficientă
- Complexitate (ex. un polinom de grad foarte înalt)

- **Exemplu:** Vizualizarea unui polinom de grad înalt care se potrivește perfect punctelor zgomotoase dar ratează tendința generală

# 7: REGULARIZARE LASSO & RIDGE

- **Scopul regularizării:** Controlul complexității modelului, reducerea varianței
- **Regularizare Ridge (L2):**
  - Aduagă un termen penalizant ( $\lambda^2$ )
  - Micșorează coeficienții dar rareori îi reduce complet la zero
- **Regularizare Lasso (L1):**
  - Aduagă un termen penalizant ( $|\lambda|$ )
  - Poate reduce unii coeficienți la zero (selecție de caracteristici)
- **Ajustarea Hiperparametrilor:**
  - Alegerea lui  $\lambda$  (forța regularizării)
- **Exemplu:** Demonstrați cum același set de date poate duce la valori diferite ale coeficienților cu Lasso față de Ridge

# 8: METRICI DE EVALUARE

- **Scorul  $R^2$  (Coeficientul de determinare):**

- Măsoară proporția de variație explicată de model
- Se situează între 0 și 1 (mai mare este mai bine)

- **Eroarea pătratică medie (MSE):**

- Media pătratelor diferențelor dintre predicții și valorile reale
- Sensibilă la valori extreme (outliers)

- **Eroarea absolută medie (MAE):**

- Media diferențelor absolute
- Mai puțin sensibilă la outliers decât MSE

- **Exemplu:** Comparați  $R^2$ , MSE și MAE pe un set mic de date cu prețuri de locuințe prezise vs. reale

- Articol



# 9: APLICAȚII ÎN LUMEA REALĂ

- **Exemple:**

- **Finanțe:** Estimarea prețului acțiunilor sau a riscului
- **Sănătate:** Prezicerea ratelor de readmisie la spital sau a costurilor medicale
- **Marketing:** Prognoza vânzărilor sau a valorii pe viață a clientului
- **Inginerie:** Prognoza cererii, controlul calității

- **Idei principale:**

- Selectați modelele în funcție de forma și complexitatea datelor
- Validați întotdeauna cu metrici corespunzătoare

- **Exemplu:** 0 scurtă demonstrație a folosirii unui set de date real (ex. cheltuieli de publicitate online vs. vânzări) pentru a construi un model de regresie

# 10: K-NEAREST NEIGHBORS (KNN) PENTRU REGRESIE

## ● Concept

- În mod de regresie, kNN prezice o valoare prin **media** valorilor țintă ale celor (k) vecini cei mai apropiați din spațiul caracteristicilor.
- Metricile de distanță (ex. Euclidiană) determină care vecini sunt „cei mai apropiați”.

## ● Pași cheie

1. Alegeți (k) (numărul de vecini).
2. Calculați distanțele dintre eșantionul nou și toate eșantioanele de antrenament.
3. Selectați cei (k) cei mai apropiați puncte.
4. **Faceți media** valorilor țintă pentru a obține predicția.

# 11: K-NEAREST NEIGHBORS (KNN) PENTRU REGRESIE

- **Avantaje**

- Simplu și intuitiv.
- Funcționează bine cu relații neliniare.
- Fără fază explicită de antrenare (învățare „leneșă”).

- **Dezavantaje**

- Poate fi lent pentru seturi de date mari (calcul de distanță).
- Performanța depinde mult de alegerea lui  $k$  și de metrica de distanță.
- Sensibil la scalarea caracteristicilor și la caracteristicile irelevante.

- **Exemplu:** Prezicerea prețurilor locuințelor pe baza prețurilor celor mai apropiate locuințe în spațiul caracteristicilor (ex. dimensiune similară, locație etc.) prin media prețurilor

- **Articol**

# 12: CONCLUZII & ÎNTREBĂRI

## ● Concluzii:

- Regresia prezice valori continue
- Abordările Liniară, Polinomială și bazate pe Arbori au fiecare avantaje/dezavantaje
- Supraadaptarea poate fi gestionată prin regularizare (Lasso/Ridge)
- Modelele se evaluează cu ajutorul  $R^2$ , MSE și MAE

## ● Note Finale:

- Încurajați experimentarea cu diverse modele
- Subliniați importanța validării corespunzătoare

## ● Întrebări?