Mašinsko učenje. Regresija.

Danijela Petrović

May 17, 2016

Uvod

- Problem **predviđanja** vrednosti *neprekidnog atributa* neke instance na osnovu vrednosti njenih drugih atributa.

Uvod

- Problem **predviđanja** vrednosti *neprekidnog atributa* neke instance na osnovu vrednosti njenih drugih atributa.
- Aproksimacija neprekidne ciljne funkcije

- Pretpostavlja se linearna veza između **promenljive koja se predviđa** i **datih promenljivih (prediktora)**

Linearna regresija

- Pretpostavlja se linearna veza između **promenljive koja se predviđa** i **datih promenljivih (prediktora)**
- $Y = X\omega_1 + \omega_0$

Linearna regresija

- Pretpostavlja se linearna veza između **promenljive koja se predviđa** i **datih promenljivih (prediktora)**
- $Y = X\omega_1 + \omega_0$
- Y – *slučajna promenljiva* (tražena promenljiva)

Linearna regresija

- Pretpostavlja se linearna veza između **promenljive koja se predviđa** i **datih promenljivih (prediktora)**
- $Y = X\omega_1 + \omega_0$
- Y – *slučajna promenljiva* (tražena promenljiva)
- X – *poznati prediktori* (poznate promenljive)

Linearna regresija

- Pretpostavlja se linearna veza između **promenljive koja se predviđa** i **datih promenljivih (prediktora)**
- $Y = X\omega_1 + \omega_0$
- Y – *slučajna promenljiva* (tražena promenljiva)
- X – *poznati prediktori* (poznate promenljive)
- $\omega = (\omega_1, \omega_0)$ – *nepoznati vektor*

Linearna regresija

- Pretpostavlja se linearna veza između **promenljive koja se predviđa** i **datih promenljivih (prediktora)**
- $Y = X\omega_1 + \omega_0$
- Y – *slučajna promenljiva* (tražena promenljiva)
- X – *poznati prediktori* (poznate promenljive)
- $\omega = (\omega_1, \omega_0)$ – *nepoznati vektor*
- Zadatak regresije je da odredimo ω (na osnovu trening podataka, opažanjima iz iskustva)

Linearna regresija

- Pretpostavlja se linearna veza između **promenljive koja se predviđa** i **datih promenljivih (prediktora)**
- $Y = X\omega_1 + \omega_0$
- Y – *slučajna promenljiva* (tražena promenljiva)
- X – *poznati prediktori* (poznate promenljive)
- $\omega = (\omega_1, \omega_0)$ – *nepoznati vektor*
- Zadatak regresije je da odredimo ω (na osnovu trening podataka, opažanjima iz iskustva)
- Greška se ne modelira

Linearna regresija

- $\omega = (X^T X + \lambda I)^{-1} X^T Y$
  često se za λ uzima vrednost 0

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Ispitivanje kvaliteta linearne regresije

- formalno – korišćenjem statističkih testova

Ispitivanje kvaliteta linearne regresije

- formalno – korišćenjem statističkih testova
- neformalno – pomoću dijagrama

Ispitivanje kvaliteta linearne regresije

- formalno – korišćenjem statističkih testova
- neformalno – pomoću dijagrama
- korišćenjem **reziduala** – $r_i = y_i - \omega x_i$

Ispitivanje kvaliteta linearne regresije

- formalno – korišćenjem statističkih testova
- neformalno – pomoću dijagrama
- korišćenjem **reziduala** – $r_i = y_i - \omega x_i$
- osnovna mera kvaliteta je *srednjekvadratna greška*:

$$E(Y, X\omega) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \omega x_i)^2$$

Ispitivanje kvaliteta linearne regresije

- formalno – korišćenjem statističkih testova
- neformalno – pomoću dijagrama
- korišćenjem **reziduala** – $r_i = y_i - \omega x_i$
- osnovna mera kvaliteta je *srednjekvadratna greška*:

$$E(Y, X\omega) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \omega x_i)^2$$

- Upotrebljava se još i *koeficijent determinacije:*

$$r^2(Y, X\omega) = 1 - \frac{\sum_{i=1}^{n}(y_i - \omega x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

pri čemu je $\bar{y}$ prosek uzorka.

Ispitivanje kvaliteta linearne regresije

- Ako je vrednost koeficijenta determinacije 1 – postoji potpuno podudaranje stvarnih i predviđenih verdnosti.

Ispitivanje kvaliteta linearne regresije

- Ako je vrednost koeficijenta determinacije 1 – postoji potpuno podudaranje stvarnih i predviđenih verdnosti.
- Što je vrednost koeficijenta determinacije manja, to je poklapanje lošije.

Primer

Instrument meri brzinu tela u padu. Izmerena brzina je 2m/s u polaznom trenutku, 4 dve desetinke kasnije, a 6.9 pola sekunde kasnije (u odnosu na polazni trenutak). Linearnom regresijom odrediti model koji predvidja brzinu tela u buducnosti i proceniti brzinu posle jedne i posle dve sekunde. Na osnovu modela proceniti ubrzanje sa koje Zemljina teza uzrokuje u kretanju tela.

Primer

Jedne nedelje januara, u ponedeljak, utorak i petak u podne izmerene su temperature -2, 0 i 1 stepen. Linearnom regresijom proceniti temperaturu u sredu i četvrtak u podne. Koliki je koeficijent determinacije za dobijeni linearni model?

Primer

Za količine katalizatora od 0,1 i 2 grama, izmerene su brzine hemijske reakcije od 5, 6 i 1 sekunde. Pomoću koeficijenta korelacije oceniti kvalitet linearnog modela t=6-2m dobijenog linearnom regresijom iz datih podataka. Odrediti koeficijent deteminacije. Šta možemo reći na osnovu njegove vrednosti.

Primer

Vrednost evra 3. juna je 100 dinara, 4. juna je 101 dinar, a 5. juna je 105 dinara. Pomoću linearne regresije predvideti vrednost evra 6., 7. i 8. juna. Stvarne vrednosti tih dana su bile 105, 106 i 107. Kolika je srednjekvadratna greška tih predviđanja?

Primer

Usled poremećaja na tržištu prirodnog gasa, došlo je do brzog rasta cena. U ponedeljak, u utorak i u četvrtak u podne cene kubnog metra su bile 3, 5 i 6 dolara. Proceniti cenu gasa u sredu u 04:00h i petak u 18:00h (primenom linearnog regresionog modela).

Primer

Linearnom regresijom odrediti gustinu tela ako se zna da je telo zapremine $5m^3$ mase $2kg$, telo zapremine $10m^3$ mase $4kg$, a telo zapremine $18m^3$ mase $7.2kg$.

Primer

Telo se krece po putu konstanantnom brzinom. Nakon jedne sekunde telo je prešlo 6m od starta, nakon 2s 8m, a nakon 3s 10m. Koriteći lineranu regresiju odrediti brzinu tela i na kojoj razdaljini od starta je bilo telo u početnom trenutku.

Primer

U toku dana praćena je temperatura vazduha. U 8:00 ujutru je bilo 15 stepeni, a u 10:00 je bilo 18 stepeni. Linearnom regresijom odrediti model koji predvidja temperaturu u budućnosti i proceniti temperaturu u 12:00 i 14:00.