Regresija z Gaussovimi procesi

Sara Kovačič

Mentor: doc. dr. Aljoša Peperko

Večrazsežna normalna porazdelitev

Slučajni vektor $X = [X_1, X_2, \ldots, X_n]$ ima večrazsežno normalno porazdelitev s povprečjem (matematičnim upanjem) $\mu \in \mathbb{R}^n$ in kovariančno matriko $\Sigma \in \mathbb{R}^{n \times n}$, če je njena funkcija gostote enaka:

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} \cdot \left|\Sigma^{1/2}\right|} exp\left(-\frac{1}{2}(x-\mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)\right).$$

Pišemo $X \sim N(\mu, \Sigma)$.

Strojno učenje

- Strojno učenje je področje umetne inteligence, ki se ukvarja z razvojem tehnik, ki računalnikom (oz. strojem) omogočajo, da se lahko učijo.
- Je metoda za kreiranje računalniških programov na podlagi podatkov (vzorcev).
- Vhodni podatki: x
- Izhodni podatki: y

Enostavna linearna regresija

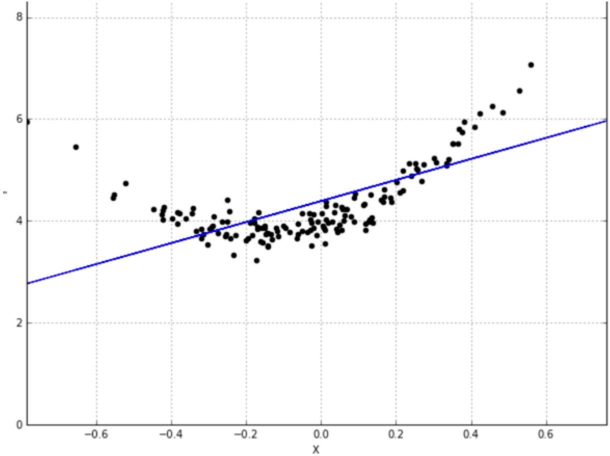
$y = f(x) + \varepsilon$, kjer je ε slučajno odstopanje od linearne zveze.

- $E(\varepsilon) = 0$.

Iščemo torej parametra β_0 in β_1 za $y = \beta_0 + \beta_1 \cdot x + \varepsilon$.

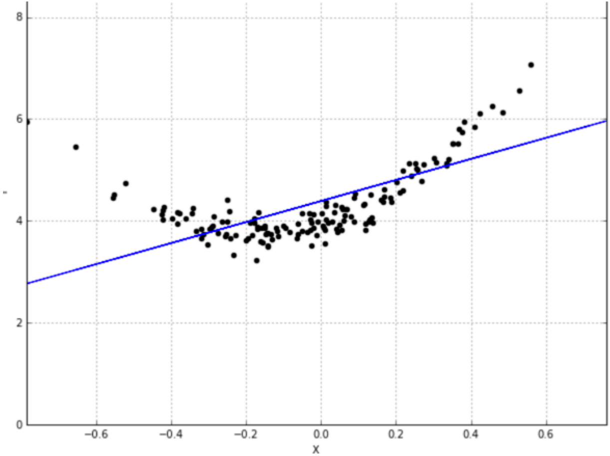
- Bayesova linearna regresija: porazdelitev nad parametri ki se posodabljajo ob vsakem opazovanju novih točk.
- Regresija z GP: neparametričen pristop, saj najde porazdelitev nad možnimi funkcijami $f(x)$, ki so skladne z opazovanimi podatki.

Problem linearne regresije



$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \varepsilon$$

Problem linearne regresije



$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \varepsilon$$

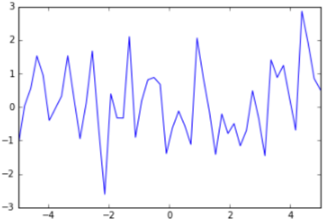
Problem linearne regresije

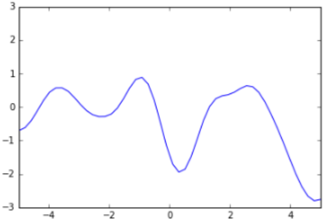
Kaj če ne želim (ne znam) v naprej določiti števila parametrov?

- Želimo upoštevati vse možne funkcije, ki se ujemajo z našimi podatki, ne glede na to, koliko parametrov vsebujejo.
- Neparametrično = neskončno mnogo parametrov

Omejitev funkcij, ki jih želimo

- Zaloga vrednosti
- Gladkost
- Povprečna vrednost

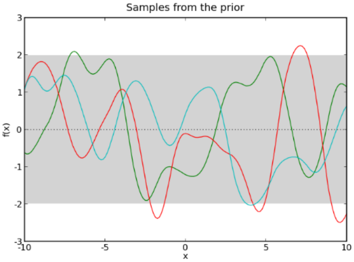


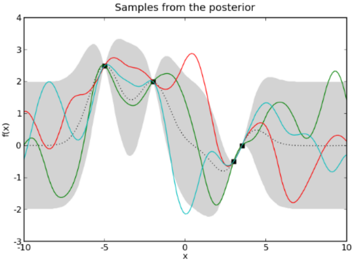


Gladkost

- Podobne vrednosti vhodnih podatkov producirajo podobne vrednosti izhodnih
- Kovariančna matrika (kovariančna funkcija)

Prior in posterior





Definicija

Slučajni proces je zaporedje slučajnih spremenljivk $(X_t)_{t \geq 0}$.

Definicija

Slučajni proces $(X_t)_{t \in T}$ je Gaussov, če je za katerokoli končno podmnožico $F \subset T$, slučajni vektor $X_F := (X_t)_{t \in F}$ (večrazsežno) normalno porazdeljen.

Drugače: Slučajni proces je Gaussov, če je za vsak vektor neodvisnih spremenljivk x, vrednost funckije $f(x)$ porazdeljena po normalni (Gaussovi) porazdelitvi.

Gaussov proces

Gaussov proces $f(x)$ je določen s funkcijo matematičnega upanja $m(x)$ in kovariančno funkcijo $k(x, x')$, kjer sta:

- $m(x) = E[f(x)]$,
- $k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]$.

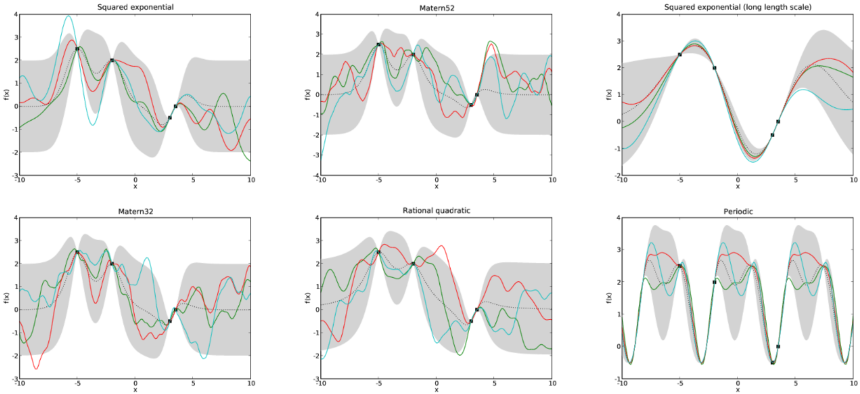
Pišemo: $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$.

Kovariančna funkcija

Vrednost kovariančne funkcije $k(x_i, x_j)$ izraža korelacijo med posameznima izhodoma $f(x_i)$ in $f(x_j)$ modela, obravnavana kot dve medsebojno povezani naključni spremenljivki.

- $k(x_i, x_j) = E[(f(x_i) - m(x_i))(f(x_j) - m(x_j))]$.

Primerjava različnih kovarianc



GP model

- **Predhodno znanje** (ang. *prior*) odraža mnenje o preslikavi med vhodi in izhodi. Običajno predpostavlja gladkost.
- **Posteriorno znanje** (ang. *posterior*) dobimo, ko v model vključimo še končno število vhodno-izhodnih parov (x_i, y_i). Dobimo posteriorno porazdelitev za predikcijo modela.
- **Vhod** v GP model: posamezne vrednosti neodvisnih spremenljivk, zbrane v vhodnem vektorju x
- **Izhod** iz GP modela: verjetnostna porazdelitev izhodne vrednosti $f(x)$ pri danem vhodnem vektorju

Delo v nadaljevanju

- Izbira kovariančne funkcije je ključnega pomena za delovanje GP modela.
- Empirični del diplomske naloge