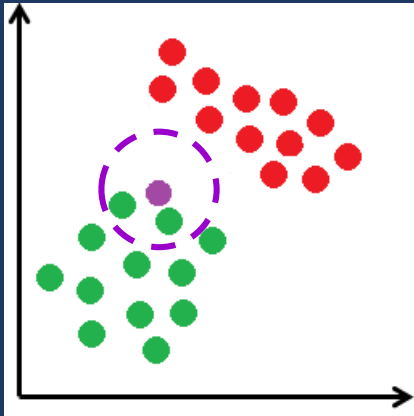


Generativni modeli

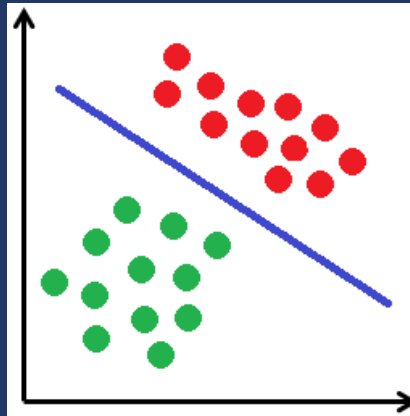
Tipovi klasifikatora

Bazirani na instancama



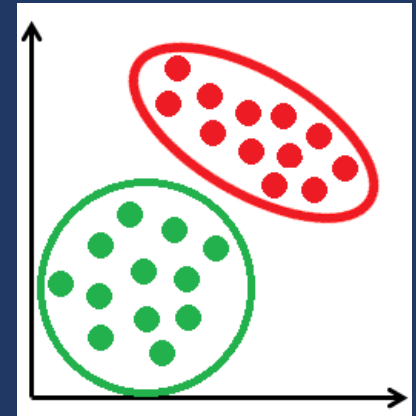
- KNN
- Koriste opservacije direktno (bez formiranja modela)
- Ne uvode nikakve pretpostavke o strukturi podataka

Diskriminativni



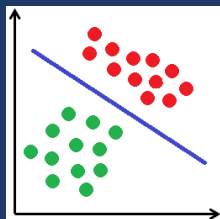
- Logistička regresija, SVM, Boosting
- Direktno estimiraju granicu odluke

Generativni



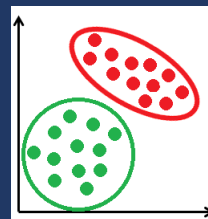
- Naivni Bajes, model Gausovih mešavina
- Modeluju kako izgledaju podaci svake klase

Diskriminativni



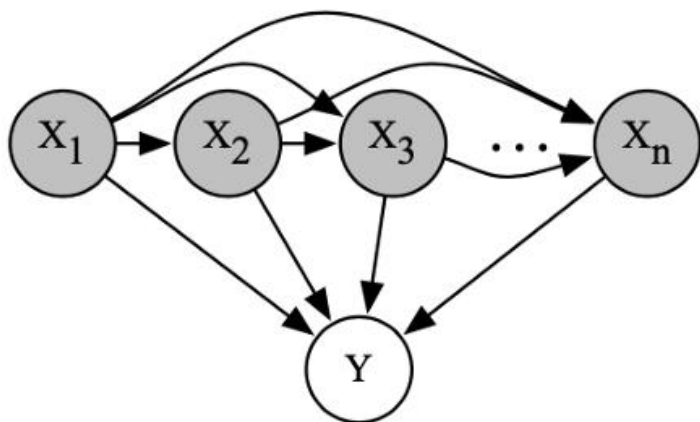
- Direktno učimo $P(y|x)$:
 - Prepostaviti formu $P(y|x)$
 - Na primer: u logističkoj regresiji prepostavljamo sigmoid
- Ne možemo generisati podatke jer ne znamo $P(x)$

Generativni

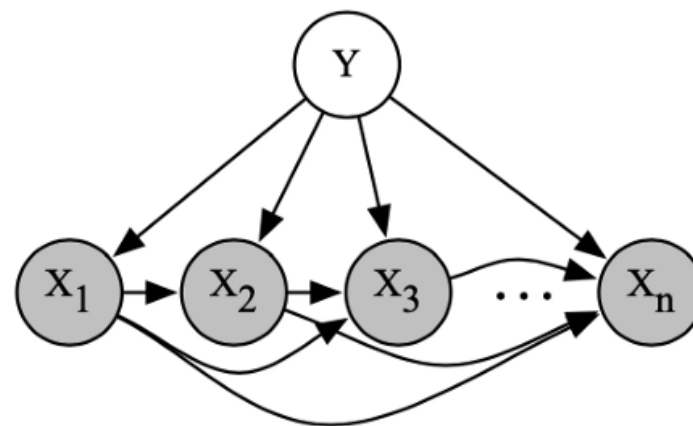


- Određujemo $P(x|y)$ i $P(y)$
 - Prepostaviti formu $P(x|y)$ i $P(y)$
 - Odvojeno modelovati kako izgleda x jedne klase i druge klase
 - $P(y)$ - apriorna verovatnoća y
 - Bajesovim pravilom (indirektno) određujemo $P(y|x)$
- Možemo generisati primere
$$P(x) = \sum_y P(y)P(x|y)$$

Diskriminativni

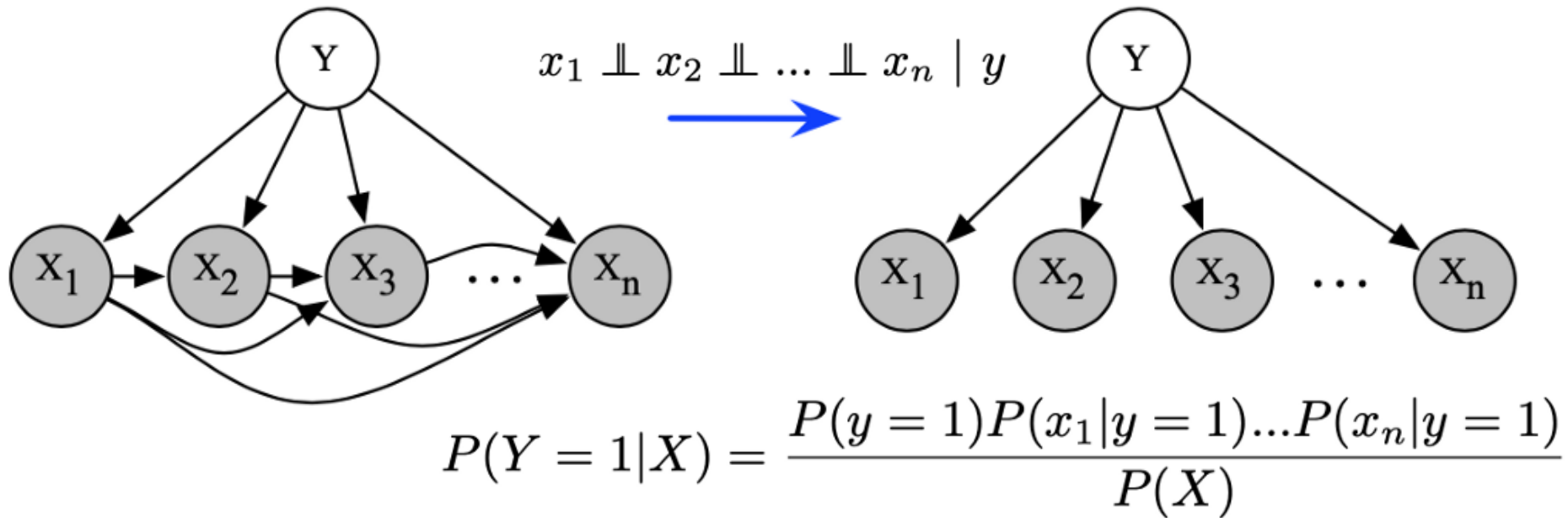


Generativni



Naivna pretpostavka

$$P(Y|X) = P(y)P(x_1|y)P(x_2|y, x_1)...P(x_n|y, x_1...x_{n-1})$$



Poređenje: tačnost

- Generativni modeli su lošiji kada pretpostavka o uslovnoj nezavisnosti nije zadovoljena
- x_1 : broj pojave *bank*
- x_2 : broj pojave *account*
- Nezavisno da li se radi o *spam* mailu, ove dve reči se uvek pojavljuju zajedno ($x_1 = x_2$)
- Naive Bayes: $P(x_1|y) = P(x_2|y)$ - duplira dokaz
- Logistička regresija: može da podesi θ_1 ili θ_2 na 0

Poređenje: tačnost

- Generativni modeli uvode snažnije pretpostavke
- Potrebno je napraviti pretpostavke o raspodeli verovatnoće (a one mogu biti pogrešne)
- Ali, zato im treba manje podataka za trening
- Pokazalo se da, pogotovo u slučaju malog skupa podataka, NB često pokazuje bolje performanse od sofisticiranijih alternativa

[P. Domingos and M. Pazzani, "On the optimality of the simple bayesian classifier under zero-one loss," Machine learning, vol. 29, no. 2-3, pp. 103–130, 1997.]

- Ignoriše interakcije među atributima pa mu generalno treba manje podataka u odnosu na npr. logističku regresiju
- Manje verovatan da overfituje u slučaju male količine podataka

Poređenje: neodstajuće vrednosti

- Generativni modeli mogu da rade sa nedostajućim vrednostima
- Diskriminativni modeli (u opštem slučaju) ne mogu (zahtevaju sve ulaze x)

Poređenje: primena

- Generativni modeli mogu da generišu podatke (sa istim svojstvima poput podataka na kojima je trenirano)
- Da li je potrebno poznavati zavisnosti između atributa kako bi se predvidela ciljna promenljiva?
- Diskriminativni model mogu da rešavaju klasifikacione probleme
- Generativni modeli imaju mnogo drugih primena:
 - Uzorkovanje
 - Generisanje podataka (npr. *Generative Adversarial Networks* za generisanje novih slika)