

3. (a) Svaki algoritam strojnog učenja sastoji se od tri osnovne komponente. Identificirajte te komponente kod naivnog Bayesovog klasifikatora.

**Model:**

$$h(x_1, \dots, x_n) = \underset{j}{\operatorname{argmax}} P(C_j) \prod_{k=1}^n P(x_k | C_j)$$

**Funkcija gubitka:**

Općenitiji slučaj, kada gubitci ne moraju biti jednoliki

Funkcija gubitka je vjerojatnost pogrešne klasifikacije primjera. U slučaju da gubitci pri pogrešnoj klasifikaciji nisu jednoliki, koristimo matricu gubitka  $[L_{kj}]$ , gdje je  $L_{kj}$  gubitak prilikom pogrešne klasifikacije primjera  $\mathbf{x}$  iz klase  $C_k$  u klasu  $C_j$ .

Očekivana vrijednost funkcije gubitka  $L$  naziva se funkcija rizika i definirana je kao:

$$\mathbb{E}[L] = \sum_{k=1}^K \sum_{j=1}^K \int_{\mathbf{x} \in \mathcal{R}_j} L_{kj} p(\mathbf{x}, C_k) d\mathbf{x}$$

Područje integracije obuhvaća sve primjere koje klasifikator klasificira kao  $C_j$ , odnosno  $\mathcal{R}_j = \{ \mathbf{x} \in \mathcal{X} \mid h_j(\mathbf{x}) = 1 \}$ .

**Optimizacijski postupak:**

Minimizacija očekivanog gubitka  $\mathbb{E}[L]$  svodi se na minimiziranje očekivanog rizika:

$$R(C_j | \mathbf{x}) = \sum_{k=1}^K L_{kj} P(C_k | \mathbf{x})$$

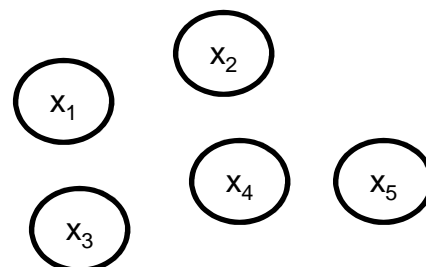
Za optimalna klasifikacija u smislu minimizacije rizika, funkciju hipoteze možemo definirati kao:

$$h(\mathbf{x}) = \underset{C_k}{\operatorname{argmin}} R(C_k | \mathbf{x})$$

(b) Faktorizirajte zajedničku vjerojatnost  $P(x_1, \dots, x_5)$  na tri različita načina. Koje pretpostavke o nezavisnosti varijabli proizlaze iz tih faktorizacija? Nacrtajte odgovarajuće Bayesove mreže.

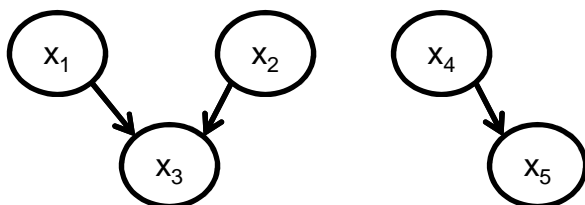
1)  $P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2)P(x_3)P(x_4)P(x_5)$

Pretpostavka: sve varijable su nezavisne.



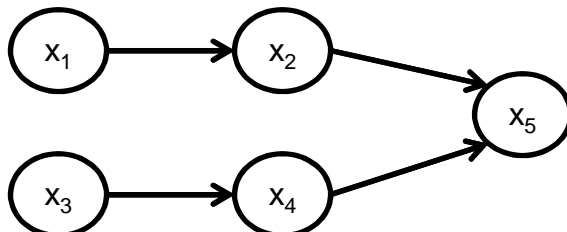
2)  $P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2)P(x_3|x_1, x_2)P(x_4)P(x_5|x_4)$

Pretpostavka: varijabla  $x_3$  je zavisna o  $x_1$  i  $x_2$ , varijabla  $x_5$  je zavisna o  $x_4$ .



3)  $P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2|x_1)P(x_3)P(x_4|x_3)P(x_5|x_2, x_4)$

Pretpostavka: varijabla  $x_2$  je zavisna o  $x_1$ , varijabla  $x_4$  je zavisna o  $x_3$ , varijabla  $x_5$  je zavisna o  $x_2$  i  $x_4$ .



(c) Pretpostavite da se radi o binarnoj klasifikaciji i da su dvije značajke binarne, a tri su ternarne. Izračunajte broj parametara svakog od tri modela.

1)  $P(C_i)P(x_1|C_i)P(x_2|C_i)P(x_3|C_i)P(x_4|C_i)P(x_5|C_i)$

$$(2-1) + 2 * 2 * (2-1) + 3 * 2 * (3-1) = 17$$

$x_1, x_2$  - binarni

$x_3, x_4, x_5$  - ternarni

2)  $P(C_i)P(x_1|C_i)P(x_2|C_i)P(x_3|x_1, x_2, C_i)P(x_4|C_i)P(x_5|x_4, C_i)$

$$(2-1) + 2 * 2 * (2-1) + 2 * 2^2 * (3-1) + 2 * (3-1) + 2 * 3 * (3-1) = 37$$

3)  $P(C_i)P(x_1|C_i)P(x_2|x_1, C_i)P(x_3|C_i)P(x_4|x_3, C_i)P(x_5|x_2, x_4, C_i)$

$$(2-1) + 2 * (2-1) + 2 * 2 * (2-1) + 2 * (3-1) + 2 * 3 * (3-1) + 2 * 3 * 2 * (3-1) = 47$$

(d) Osmislili ste svoj potpuno nov algoritam za konstrukciju polunaivnog Bayesovog klasifikatora koji ste nazvali ASJ (*A-star Selection and Joining*). ASJ je napredna verzija algoritma FSSJ. Isto kao i algoritam FSSJ, algoritam ASJ čvorove spaja u superčvorove, ali, za razliku od algoritma FSSJ, koji koristi pohlepno pretraživanje, algoritam ASJ za pretraživanje koristi algoritam  $A^*$ . Pritom se za heuristiku koristiti empirijska pogreška, tj. u svakoj se iteraciji odabire ona operacija koja smanjuje pogrešku na skupu za učenje.

Je li algoritam ASJ, kako je ovdje opisan, dobar? Postoje li nekakvi nedostaci ili problemi koje možemo očekivati? Obrazložite odgovore. Predložite poboljšanja.

Prvo, ako se zbilja u svakoj iteraciji odabire operacija sa manjom heuristikom, onda to nije algoritam  $A^*$ , jer  $A^*$  uzima u obzir i heuristiku i trenutnu cijenu stanja u kojem se nalazimo. To je običan pohlepni algoritam, poput FSSJ, samo sa drugom heuristikom.

Ali dobro, nastavimo dalje... ASJ, kakav je ovdje napisan, nije dobar, iz razloga što mu ne valja heuristika. Najmanju pogrešku na skupu za učenje uvijek će imati združena vjerojatnost, tako da će navedeni ASJ uvijek težiti potpuno nefaktoriziranom modelu (superčvoru sa svim varijablama), a takav model bi trebao imati veliku pogrešku generalizacije. Poboljšanja: za heuristiku uzeti zbroj pogreške na skupu za učenje i pogreške na skupu za provjeru.