

# Notițe Seminar 7

December 22, 2019

**Intro:** Seminarul trecut am început să discutăm despre algoritmi Bayes naiv și Bayes corelat. Acum vom continua discuția despre ei cu două mari noi chestiuni:

- rata medie a erorilor pentru NB și JB
- extensia lui NB pentru lucrul cu atribute continue (la intrare)

## 1 Rata medie a erorilor

### 1.1 Remember 1

Regulile de decizie finale pentru NB și JB sunt:

$$v_{NB} = \arg \max_{v_j \in V} \underbrace{P(v_j)P(a_1|v_j) \dots P(a_n|v_j)}_{P_{NB}(a_1, \dots, a_n, v_j)}$$

$$v_{JB} = \arg \max_{v_j \in V} \underbrace{P(v_j)P(a_1, \dots, a_n|v_j)}_{P_{JB}(a_1, \dots, a_n, v_j) = P(a_1, \dots, a_n, v_j)}$$

Notăția  $P_{NB}$  indică faptul că NB folosește distribuția de probabilitate în care folosim presupunerea de independență condițională a atributelor de intrare față de ieșire.

Faptul că  $P_{JB} = P$  ( $P$  - distribuția reală de probabilitate) nu ar trebui să fie o surpriză, pentru că JB nu face nicio presupunere neconformă cu realitatea.

## 1.2 Remember 2

$$E[X] = \sum_{x \in X} xP(X = x)$$

$$E[f(X)] = \sum_{x \in X} f(x)P(X = x)$$

$$E[XY] = \sum_{x \in X, y \in Y} xyP(X = x, Y = y)$$

$$E[f(X, Y)] = \sum_{x \in X, y \in Y} f(x, y)P(X = x, Y = y)$$

## 1.3 Remember 3

Funcția **indicator**:  $I : \{\text{True}, \text{False}\} \rightarrow \{0, 1\}$ ,  $I(\text{True}) = 1$ ,  $I(\text{False}) = 0$ .

## 1.4 Rata medie a erorilor: $E_P[I(v_{NB/JB}(A_1, \dots, A_n) \neq V)]$

$$\begin{aligned} E \underbrace{P}_{\text{distribuția reală}} \underbrace{[I(v_{NB/JB}(A_1, \dots, A_n) \neq V)]}_{f(A_1, \dots, A_n, V)} &\stackrel{\text{def.}}{=} \\ \stackrel{\text{def.}}{=} \sum_{a_1 \in A_1, \dots, a_n \in A_n, v_j \in V} \underbrace{I(v_{NB/JB}(A_1 = a_1, \dots, A_n = a_n) \neq v_j)}_{f(a_1, \dots, a_n, v_j)} \underbrace{P(A_1 = a_1, \dots, A_n = a_n, V = v_j)}_{\text{distribuția reală}} \end{aligned}$$

Exercițiu: vezi ex. 28/pag. 414. Vezi rezolvarea din slide-uri: <https://profs.info.uaic.ro/~ciortuz/ML.ex-book/SLIDES/ML.ex-book.SLIDES.Bayes.pdf> (începând cu slide-ul #48).

## 1.5 Surse de erori pentru NB și JB

Bayes Naiv are două surse de erori:

- votul majoritar (argmax)
- presupunerea de independență condițională

Bayes Corelat are o singură sursă de erori:

- votul majoritar (argmax)

Dacă datele sunt consistente, atunci sursa de eroare legată de votul majoritar dispare.

Astfel, pentru un set de date consistent, Bayes Corelat nu are nicio sursă de erori, deci, rata medie a erorii pentru el este 0.

## 2 Extensie Bayes Naiv - attribute continue la intrare

### 2.1 Intro 1

Semnul  $\sim$  este folosit în contextul PS.

Dacă  $X \sim \text{Distribuție}(\text{parametri})$ , atunci  $X$  este o variabilă aleatoare care urmează distribuția  $\text{Distribuție}(\text{parametri})$ .

### 2.2 Intro 2

- Dacă  $X \sim \text{Bernoulli}(\theta)$ , atunci  $X : \begin{pmatrix} 0 & 1 \\ 1 - \theta & \theta \end{pmatrix}$ .

$p$  este parametrul pentru distribuția Bernoulli.

Funcția masă de probabilitate (pmf) pentru  $X$  este

$$p(x) = P(X = x) = \text{Bernoulli}(x; \theta) = \begin{cases} \theta & x = 1 \\ 1 - \theta & x = 0 \end{cases}.$$

MLE pentru  $\theta$ :

Exemplu:

Fie setul de date  $D = (1, 0, 1, 0, 1)$ . Atunci:

$$\theta_{\text{MLE}} = \frac{3}{5}$$

- Dacă  $X \sim \text{Categorială}(p_{v_1}, \dots, p_{v_n})$ , atunci  $X : \begin{pmatrix} v_1 & \dots & v_n \\ p_{v_1} & \dots & p_{v_n} \end{pmatrix}$ .

$p_{v_1}, \dots, p_{v_n}$  sunt parametrii pentru distribuția Categorială.

Funcția masă de probabilitate (pmf) pentru  $X$  este

$$p(x) = P(X = x) = \text{Categoricală}(x; p_{v_1}, \dots, p_{v_n}) = \begin{cases} p_{v_1} & x = v_1 \\ p_{v_2} & x = v_2 \\ \vdots & \\ p_{v_n} & x = v_n \end{cases}.$$

MLE pentru  $p_{v_1}, \dots, p_{v_n}$ :

Exemplu:

Fie setul de date  $D = (0, 0, 1, 1, 2)$ . Atunci:

$$p_{0;\text{MLE}} = \frac{2}{5}$$

$$p_{1;\text{MLE}} = \frac{2}{5}$$

$$p_{2;\text{MLE}} = \frac{1}{5}$$

## 2.3 Intro 3: Variabile aleatoare (VA) continue

**Intro:** Dacă mulțimea de valori pentru o variabilă aleatoare discretă era o mulțime numărabilă, mulțimea de valori pentru o variabilă aleatoare continuă va fi  $\mathbb{R}$  sau un interval din  $\mathbb{R}$ . Un exemplu de valoare aleatoare continuă este înălțimea oamenilor.

<b>X - VA discretă</b>	<b>X - VA continuă</b>
Val(X) - mulțime <i>discretă</i>	Val(X) - mulțime <i>continuă</i>
Exemplu de experiment din spate: Aruncăm o monedă și vedem dacă iese H sau T	Exemplu de experiment din spate: Alegem un om și îi măsurăm înălțimea
Probabilitățile asociate lui $X$ sunt descrise prin: $P(X = a) = \dots, \forall a \in \text{Val}(X)$	Probabilitățile asociate lui $X$ sunt descrise prin: $P(X \in [a, b]) = \dots, \forall [a, b] \subseteq \text{Val}(X)$
Există funcția masă de probabilitate (pmf) $p$ prin care putem calcula orice probabilitate de tipul $P(X = a)$ : $P(X = a) = p(a)$	Există funcția densitate de probabilitate (pdf) $p$ prin care putem calcula orice probabilitate de tipul $P(X \in [a, b])$ : $P(X \in [a, b]) = \int_{t=a}^b p(t)dt$ Obs.: $P(X = a) = P(X \in [a, a]) = \int_{t=a}^a p(t)dt = 0!!!$
Cum verificăm dacă $p$ este pmf: <ul style="list-style-type: none"> <li>• <math>p(x) \geq 0, \forall x \in \text{Val}(X)</math></li> <li>• <math>\sum_{x \in \text{Val}(X)} p(x) = 1</math></li> </ul>	Cum verificăm dacă $p$ este pdf: <ul style="list-style-type: none"> <li>• <math>p(x) \geq 0, \forall x \in \text{Val}(X)</math></li> <li>• <math>\int_{x \in \text{Val}(X)} p(x)dx = 1</math></li> </ul>
$E[X] = \sum_{x \in \text{Val}(X)} xp(x)$	$E[X] = \int_{x \in \text{Val}(X)} xp(x)dx$
Următoarele sunt valide pentru $p$ : Var, formula lui Bayes, probabilități marginale/condiționate etc.	Următoarele sunt valide pentru $p$ : Var, formula lui Bayes, probabilități marginale/condiționate etc.

Exemplu de distribuție continuă: distribuția normală/Gaussiană.

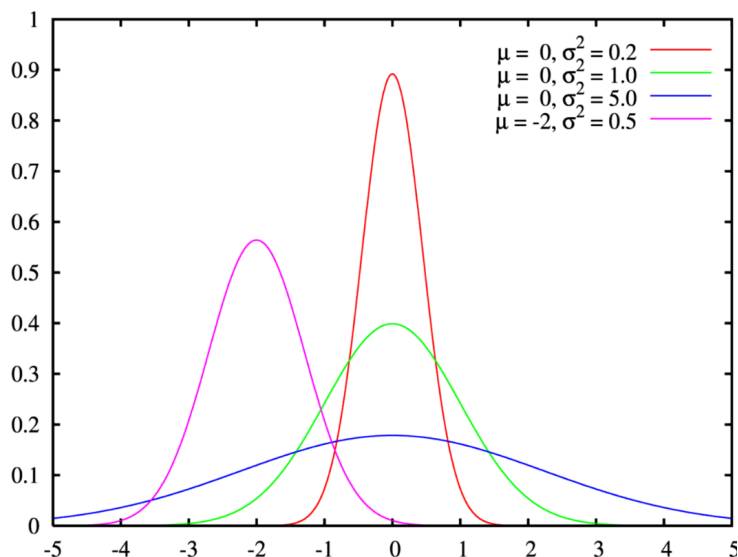
### 2.3.1 Distribuția Normală (univariată)

Dacă  $X \sim \mathcal{N}(\mu, \sigma^2)$  (adică  $X$  urmează o distribuție normală/Gaussiană), atunci  $p(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ , unde  $p$  este pdf-ul lui  $X$ .

$\mu, \sigma^2$  sunt parametrii pentru distribuția normală.

$\mu$  se mai cheamă medie, iar  $\sigma^2$  se mai cheamă varianță. Oare de ce?  
Pentru că se poate demonstra că  $E[X] = \mu$ , iar  $\text{Var}[X] = \sigma^2$ .

Graficele pentru pdf-ul  $p$  în funcție de diferite setări pentru parametrii  $\mu$  și  $\sigma^2$  sunt (pe axa orizontală aveți  $x$ , iar pe cea verticală aveți  $p(x)$ ):



(imagine preluată de aici: [https://ro.wikipedia.org/wiki/Distribu%C8%99Bia\\_Gauss#/media/Fi%C8%99ier:Normal\\_distribution\\_pdf.png](https://ro.wikipedia.org/wiki/Distribu%C8%99Bia_Gauss#/media/Fi%C8%99ier:Normal_distribution_pdf.png))

**Estimările în sensul verosimilății maxime (MLE) pentru parametrii  $\mu, \sigma^2$**

Ca de obicei (cum am mai discutat în primele seminarii pentru Bernoulli spre exemplu), în lumea reală parametrii unei distribuții de probabilitate nu sunt cunoscuți. Ceea ce cunoaștem sunt niște date (de exemplu: înălțimile unor oameni), iar din date vom estima parametrii.

Exemplu:

Fie următorul set de date:  $D = (1.5, 2, 3, 4, 5.5)$ . Atunci:

$$\begin{aligned}\mu_{\text{MLE}} &= \frac{1.5 + 2 + 3 + 4 + 5.5}{5} = 3.2 \\ \sigma_{\text{MLE}}^2 &= \frac{(1.5 - \mu_{\text{MLE}})^2 + (2 - \mu_{\text{MLE}})^2 + (3 - \mu_{\text{MLE}})^2 + (4 - \mu_{\text{MLE}})^2 + (5.5 - \mu_{\text{MLE}})^2}{5} = \\ &= \frac{(1.5 - 3.2)^2 + (2 - 3.2)^2 + (3 - 3.2)^2 + (4 - 3.2)^2 + (5.5 - 3.2)^2}{5} = 2.06\end{aligned}$$

### 2.3.2 Distribuția Normală bivariată

Fie  $X : \Omega \rightarrow \mathbb{R}^d$ . Dacă  $X \sim \mathcal{N}(\mu, \Sigma)$  (adică  $X$  urmează o distribuție normală/Gaussiană multivariată), atunci

$$p(x) = \mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)},$$

unde  $p$  este pdf-ul lui  $X$ .  $X = \begin{bmatrix} X_1 \\ \vdots \\ X_d \end{bmatrix} \in \mathbb{R}^d$

$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix} \in \mathbb{R}^d$ ,  $\Sigma \in \mathbb{R}^{d \times d}$  sunt parametrii pentru distribuția normală multivariată.

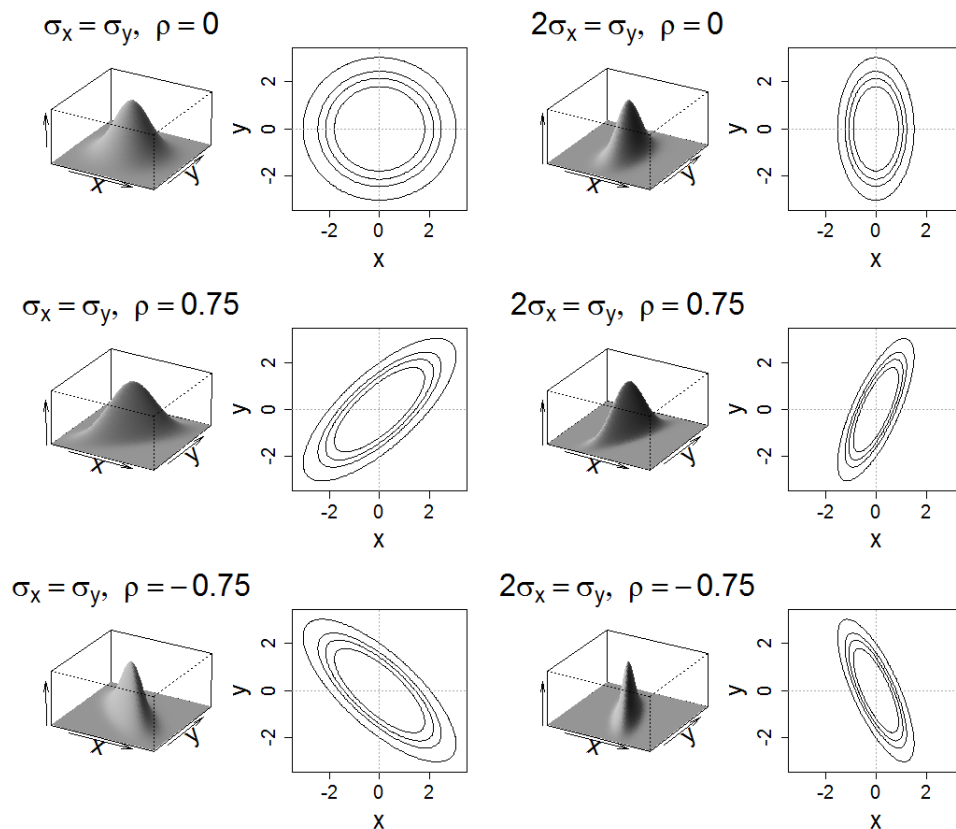
$\mu$  se mai cheamă medie, iar  $\Sigma$  se mai cheamă matrice de covarianță. Oare de ce? Pentru că se poate demonstra că  $E[X_i] = \mu_i, i \in \{1, \dots, d\}$  și  $\text{Cov}(X_i, X_j) = \Sigma_{ij}, \forall i, j \in \{1, \dots, d\}$ .

Distribuția normală bivariată este distribuția multivariată atunci când  $d = 2$ . În acest caz,  $\Sigma$  mai este scris astfel:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

unde  $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$  este numit coeficientul de corelație.

Pentru cazul bivariat, graficele pentru pdf-ul  $p$  în funcție de diferite setări pentru parametrii  $\mu$  și  $\Sigma$ :



(imagine preluată de aici: <http://www.stat.cmu.edu/~kass/KEB/figures/bivariateNormalRev.png>)

Graficele de mai sus în 3D sunt graficele pentru pdf-ul  $p$ . Graficele în 2D sunt tăieturi ale graficului în 3D cu un plan paralel cu planul de jos (adică  $Oxy$ ). În exercițiile din carte aveți de-a face cu grafice în 2D. Curbele din graficele 2D se numesc *curbe de izocontur*.

**Proprietate:** Produsul a două pdf-uri ale unor variabile normale univariate independente furnizează un pdf al unei variabile normale bivariate. Urmăriți exemplul:

Dacă  $X_1 \sim \mathcal{N}(1, 2^2)$  cu pdf-ul  $p_1$ ,  $X_2 \sim \mathcal{N}(3, 4^2)$  cu pdf-ul  $p_2$ , atunci pdf-ul  $p(x_1, x_2) = p_1(x_1)p_2(x_2)$  corespunde lui  $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2^2 & 0 \\ 0 & 4^2 \end{bmatrix}\right)$ .

**Observație** în legătură cu exemplul anterior: Din matricea  $\begin{bmatrix} 2^2 & 0 \\ 0 & 4^2 \end{bmatrix}$



observăm că  $\sigma_{12} = 0 \Rightarrow \rho = 0$ . Deci, dacă urmărim și desenul anterior, curbele de izocontur vor semăna cu desenele de pe primul rând (mai mult: vom fi pe cazul coloanei a doua pentru că  $\sigma_{X_2} > \sigma_{X_1}$ ).

## 2.4 Extensie Bayes Naiv - attribute continue la intrare

Vezi ex. 15/pag. 400.

## 2.5 Privire de final Bayes Naiv

În notițele de la seminarul trecut am scris o altă perspectivă asupra alg. NB și JB: *având un rând la testare  $(a_1, \dots, a_n)$ , gândiți-vă la rândurile  $(a_1, \dots, a_n, v_1)$ ,  $(a_1, \dots, a_n, v_2)$ , ...,  $(a_1, \dots, a_n, v_k)$ . Calculați probabilitatea fiecărui rând și returnați eticheta corespunzătoare probabilității maxime.*

Putem să punem acele probabilități într-o sumă astfel:

$$\begin{aligned} p_{\text{NB}}(a_1, \dots, a_n) &= p_{\text{NB}}(A_1 = a_1, \dots, A_n = a_n) = \\ &= p_{\text{NB}}(A_1 = a_1, \dots, A_n = a_n, V = v_1) + \dots + p_{\text{NB}}(A_1 = a_1, \dots, A_n = a_n, V = v_k) = \\ &= p_{\text{NB}}(V = v_1)p_{\text{NB}}(A_1 = a_1|V = v_1) \dots p_{\text{NB}}(A_n = a_n|V = v_1) + \dots + \\ &\quad + p_{\text{NB}}(V = v_k)p_{\text{NB}}(A_1 = a_1|V = v_k) \dots p_{\text{NB}}(A_n = a_n|V = v_k) \end{aligned}$$

Astfel, obținem pmf/pdf-ul pentru  $(A_1, \dots, A_n)$ :  $p_{\text{NB}}(a_1, \dots, a_n)$ .

Privirea de final ar fi în felul următor:

- la antrenare: scrieți  $p_{\text{NB}}(a_1, \dots, a_n)$  ca o sumă
- la testare: calculați  $p_{\text{NB}}$  pe instanța de test, termen cu termen. Returnați eticheta termenului maxim.

Exemplu:

Fie setul de date următor, unde  $X_1$  este discret, iar  $X_2$  este continuu:

$X_1$	$X_2$	Y
0	1	0
0	2	0
1	3	0
0	4	1
1	5	1

Antrenare:

$$\begin{aligned}
p_{NB}(x_1, x_2) &= p_{NB}(X_1 = x_1, X_2 = x_2, Y = 0) + p_{NB}(X_1 = x_1, X_2 = x_2, Y = 1) = \\
&= p_{NB}(Y = 0)p_{NB}(X_1 = x_1|Y = 0)p_{NB}(X_2 = x_2|Y = 0) + \\
&\quad + p_{NB}(Y = 1)p_{NB}(X_1 = x_1|Y = 1)p_{NB}(X_2 = x_2|Y = 1) \\
&= \frac{3}{5}\text{Bernoulli}(x_1; \theta = 1/3)\mathcal{N}(x_2; \mu = 2, \sigma^2 = 2/3) + \\
&\quad + \frac{2}{5}\text{Bernoulli}(x_1; \theta = 1/2)\mathcal{N}(x_2; \mu = 4.5, \sigma^2 = 0.25)
\end{aligned}$$

Testare: pe instanța (0, 6)

$$\begin{aligned}
p_{NB}(0, 6) &= \frac{3}{5}\text{Bernoulli}(0; \theta = 1/3)\mathcal{N}(6; \mu = 2, \sigma^2 = 2/3) + \\
&\quad + \frac{2}{5}\text{Bernoulli}(0; \theta = 1/2)\mathcal{N}(6; \mu = 4.5, \sigma^2 = 0.25) = \\
&= 0.0000012008 + 0.001772739
\end{aligned}$$

$$0.001772739 > 0.0000012008 \Rightarrow y_{NB} = 1$$

**Observație:** Dacă ambele atribute de intrare erau continue, nu mai înmulțeam Bernoulli cu Normală, ci Normală cu Normală. Ținând cont de această observație și folosind proprietatea și imaginea din 2.3.2 *Distribuția Normală bivariată*, puteți înțelege (mai ușor) ex. 16/pag. 402.

Vezi ex. 16/pag. 402.

## Schemă de final

1. Rata medie a erorilor
  - (a) Rata medie a erorilor
  - (b) Surse de erori pentru NB și JB
2. Extensie Bayes Naiv - attribute continue la intrare
  - (a) Remember: Variabile aleatoare discrete
    - i. Bernoulli, Categoricală
    - ii. MLE pentru Bernoulli, MLE pentru Categoricală
  - (b) Variabile aleatoare continue
    - i. Distribuția normală univariată
    - ii. MLE pentru normală univariată
    - iii. Distribuția normală bivariată
    - iv. Distribuția normală bivariată - proprietate + desen
  - (c) Privire de final Bayes Naiv