Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Naam: R. v.d. Boomgaard | Opleiding: Informatica

Collegekaartnummer: | Datum tentamen: 29/5/2015

Vak: Statistisch Redeneren | Docent: | cijfer

1. Kansrekening I

1 $\mathcal{U} = \{1, 2, \ldots, 12\}$

$$P(X=i) = \begin{cases} \frac{1}{12} & : i = 1, \ldots, 12 \\ 0 & : i \neq 1, \ldots, 12 \end{cases}$$

2 $$E(X) = \sum_x x\, P(X=x) = \sum_x x\, p_X(x)$$

Verwachting is het gemiddelde als het kansexperiment oneindig vaak wordt herhaald.

3

y	$P(Y=y)$
1	$\frac{1}{6}$
2	$\frac{1}{4}$
3	$\frac{1}{4}$
4	$\frac{1}{3}$
$y \neq 1,\ldots,4$	0

$$E(Y) = \sum_y y\, P(Y=y) = \frac{1}{6} + 2\times\frac{1}{4} + 3\times\frac{1}{4} + 4\times\frac{1}{3} = 2\frac{3}{4}$$

4 $$\mathrm{Var}(Y) = E\left((Y - EY)^2\right)$$

In diktaat is te vinden de regel:

$$E(g(X)) = \sum_x g(X)\, P(X=x)$$

nu

$$g(Y) = (Y - EY)^2$$

dus

$$\mathrm{Var}(Y) = \sum_y (y - EY)^2\, P(Y=y)$$
$$= \left(1 - 2\tfrac{3}{4}\right)^2 \times \frac{1}{6} + \left(2 - 2\tfrac{3}{4}\right)^2 \times \frac{1}{4} + \cdots$$
$$= \frac{19}{16}$$

5 $\mathrm{Var}(X) = E\left((X-EX)^2\right)$ definitie

$= E\left(X^2 - 2X\,EX + (EX)^2\right)$ uitschrijven

$= E(X^2) - E(2X\,EX) + E\left((EX)^2\right)$

$E(X+Y) = \cancel{E(X)}$

$= E(X) + E(Y)$

$= E(X^2) - 2E(X)\,E(X) + E\left((EX)^2\right)$

$E(aX) = a\,E(X)$

$= E(X^2) - 2(E(X))^2 + (EX)^2$ $\quad E(a) = a$

$= E(X^2) - (EX)^2$

BTW: ik heb op tentamens gezien dat

$E\left((X-EX)^2\right) =$

$= E\left(X^2 - (EX)^2\right)$

$= E(X^2) - (EX)^2$

dat is toch hartstikke fout!

2. Kansrekening II

1 $P(Z=2, X=3 \mid Y=1) =$

$$= \frac{P(Z=2, X=3, Y=1)}{P(Y=1)} \quad \text{def.}$$

$$= \frac{P(Z=2, X=3, Y=1)}{\sum_x \sum_z P(X=x, Y=1, Z=z)} \quad \text{marginaliseren}$$

$$= \frac{0.08}{0.05 + 0.09 + 0.13 + 0.07 + 0.06 + 0.08}$$

$$= \frac{0.08}{0.48} = \frac{1}{6}$$

2 $P(X=2, Z=1) =$ marginaliseren

$$= \sum_y P(X=2, Y=y, Z=1)$$

$$= 0{,}13 + 0{,}13 = 0{,}26$$

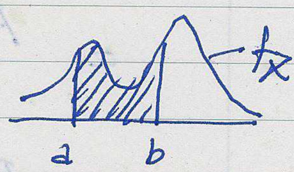
3. $P(U=u) = \sum_{i=1}^{m} P(U=u, V=v_i)$

marginaliseren (in abstracte vorm), u had het al 2x gedaan in de vorige opgaven!

3 Continue Stochasten

1 X cont. stochast als $X \in \mathbb{R}$

2 $P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$



3 $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\,dy$

N.B. mag niet x zijn!

$f_X(x) = \frac{d}{dx} F_X(x)$

4 Voorwaarde voor kansdichtheidsfunctie

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$$

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = \int_0^1 Ax^2\,dx = \left[\frac{1}{3}Ax^3\right]_0^1 = \frac{1}{3}A$$

$$\Longrightarrow \frac{1}{3}A = 1 \Longrightarrow A = 3$$

5 $E(X) = \int_0^1 x\,3x^2\,dx$

$$= \int_0^1 3x^3\,dx = \left[\frac{3}{4}x^4\right]_0^1$$

$$= \frac{3}{4}$$

UNIVERSITEIT VAN AMSTERDAM

FACULTEIT DER NATUURWETENSCHAPPEN, WISKUNDE EN INFORMATICA

Naam ______ Opleiding ______

Collegekaartnummer ______ Datum tentamen ______

Vak ______ Docent ______ cijfer

4 <u>Naive Bayes Classifier</u>

1. The JDT is veeeel te groot te weten

$$\underbrace{10}_{\text{aantal klassen}} \times \underbrace{2^{600}}_{\text{aantal mogelijke beelden}}$$

2. De 'gewone' Bayes classifier: voor alle $c' \neq c$

$$P(X_1 = x_1 \ldots X_{600} = x_{600} \mid C = c)\, P(C = c)$$

$$\geq P(X_1 = x_1 \ldots X_{600} = x_{600} \mid C = c')\, P(C = c')$$

dan classificeer als $C = c$.

Naive Bayes

$$P(X_1 = x_1 \ldots X_{600} = x_{600} \mid C = c) =$$

$$= \prod_{i=1}^{600} P(X_i = x_i \mid C = c)$$

conditionele onafhankelijkheid

Schatten van 10×600 kansen $P(X_i = x_i \mid C = c)$

3 In een beeld zijn naburige pixels verre van onafhankelijk. (Denk aan de locale structuur in beelden).

4. Voor het gemak nemen we als groepen

$$\{X_1, X_2, \ldots, X_8\} = y_1$$
$$\{X_9, \ldots, X_{16}\} = y_2$$
$$\vdots$$
$$\{X_{..}, \ldots \; X_{600}\} = y_{75}$$

N.B. y_i is een discrete stochast met uitkomstenruimte $0, \ldots, 255$ (immers 8 'bits')

$$P(X_1 = x_1 \ldots X_{600} = x_{600} \mid C = c) =$$
$$= P(y_1 = y \ldots y_{75} = y_{75} \mid C = c)$$
$$= \prod_{i=1}^{75} P(y_i = y_i \mid C = c)$$

De kansen $P(y_i = y_i \mid C = c)$ schatten door histogram te maken van uit de leervoorbeelden.

↑ 8 bins

5 Multivariate Stochasten

1. zie handouts
2. zie handouts

3. Σ : cov. matrix , μ : verwachtingsvector

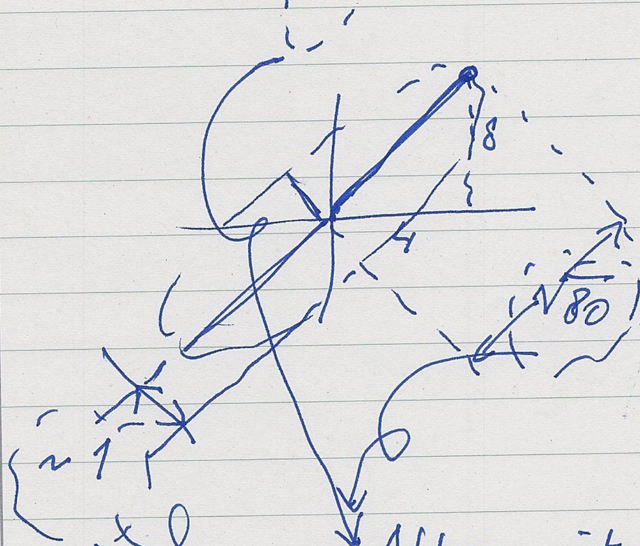
4. $\mu_X = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$ $\mu_Y = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

5. $\Sigma = U D U^T$

$$U = (u_1 \dots u_n) \qquad u_i : \text{ eigenvectoren van } \Sigma$$
$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \qquad \lambda_i : \text{ eigenwaarden van } \Sigma$$

6. $\Sigma_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (immers bijna alle punten liggen binnen cirkel met straal 3)

$$\Sigma_Y = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T$$



$\approx 3\sigma$ $\qquad \sigma = \frac{1}{3}\sqrt{80} \approx 3$

Aflezen uit de grafiek

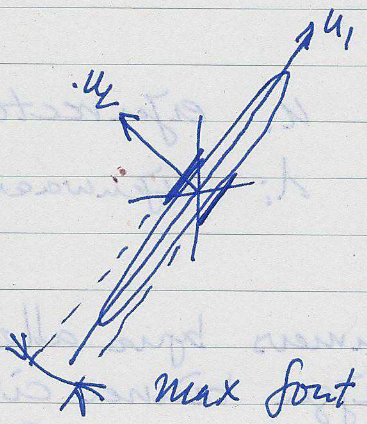
6 <u>PCA</u>

1 $y_k = u_k^T (X - m)$

$y_1 = \frac{1}{\sqrt{5}} (1 \quad 2) \begin{pmatrix} x_1 - 3 \\ x_2 - 2 \end{pmatrix}$

$= \frac{1}{\sqrt{5}} \left((x_1 - 3) + 2(x_2 - 2) \right)$

$\hat{X} = u_1 y_1 + m = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot y_1 + \begin{pmatrix} 3 \\ 2 \end{pmatrix}$

2 

max fout $\approx 3\sigma_{u_2} = 3\sqrt{0,1}$

Naam ____ Opleiding ____
Collegekaartnummer ____ Datum tentamen ____
Vak ____ Docent ____
cijfer

7. SVM

1. Linear classifier

$\vec{w}$: normaal vector op scheidslijn (hyper) plane

b : maakt dat plane niet door oorsprong hoeft. ($\propto$ met afstand van plane tot oorsprong).

2. Maximum margin

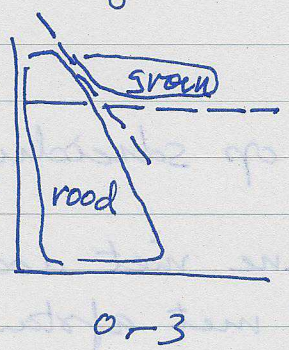
3. Kernel trick, dwz conceptuele embedding van dataset in (veel) hogere dimensie.

4. One-vs-one , One-vs-the rest

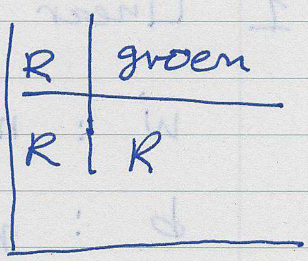
8. Classificatie

1. Op elke individuele feature as overlappen de 2 klassen behoorlijk.

2. 0 – 3 of 1 – 3



of nog eenvoudiger



N.B. Nearest neighbour is te complex!

3 Confusion matrix : zie dictaat / opgave

K : kenteken

		werkelijk K	werkelijk ¬K
classificatie	K	OK (veel)	wrong
	¬K	wrong	OK (veel)

4

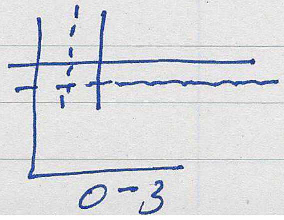
	K	¬K
K		
¬K		

dit aantal moet klein worden (tenkoste van meer in .)

Als we aannemen dat 'groen' de kentekenplaten zijn dan

∟ : oude

┆ : verbeterd.



9.

1.

2 Geef Bayes regel voor gebeurtenissen A en B

3 Wanneer zijn 2 gebeurtenissen A en B onafh.

4 Wanneer zijn 2 discrete stoch. X en Y onafh.

5. In de praktijk blijkt een data set niet lineair scheidbaar. Er zijn maar een paar leervoorbeelden die aan de verkeerde kant van het scheidingsvlak terecht komen. Wat voor SVM heb je nodig?