Tentamen: Statistisch Redeneren

docent: Rein van den Boomgaard

29 Mei 2015

1 Kansrekening I

1. Gegeven een 12 kantige dobbelsteen met daarop het standaard aantal ogen van 1 t/m 12. De dobbelsteen is zuiver. De stochast X telt het aantal ogen in 1 worp. Wat is de uitkomstenruimte voor X en wat is de kansverdeling $p_X(x)$?

2. Wat is de definitie van de verwachting van een discrete stochast (in woorden en formule). Bereken de verwachting voor stochast X uit het eerste deel.

3. Beschouw stochast Y gedefinieerd aan de hand van hetzelfde kansexperiment als hierboven beschreven (het 1 maal gooien van 12 kantige dobbelsteen als zijnde:

$$\begin{array}{lll} Y=1 & \text{als} & X=1,2 \\ Y=2 & \text{als} & X=3,4,5 \\ Y=3 & \text{als} & X=6,7,8 \\ Y=4 & \text{als} & X=9,10,11,12 \end{array}$$

   Wat is de kansverdeling van Y en wat is de verwachting van Y?

4. Wat is de definitie van de variantie van een discrete stochast? Bereken de variantie van stochast Y zoals hierboven gedefinieerd.

5. Bewijs dat voor elke discrete stochast X geldt dat: $\text{Var}(X) = E(X^2) - E^2(X)$. Geef in je bewijs aan van welke eigenschappen je gebruik maakt?

2 Kansrekening II

Gegeven de stochasten X, Y en Z. Gegeven is dat $X \in \{1,2,3\}$, $Y \in \{1,2\}$ en $Z \in \{1,2\}$. De gezamenlijke kansfunctie (joint distribution) is $P(X=x, Y=y, Z=z)$ en is geënumereerd in onderstaande tabel:

x	y	z	P(X=x, Y=y, Z=z)
1	1	1	0.05
1	1	2	0.09
1	2	1	0.12
1	2	2	0.04
2	1	1	0.13
2	1	2	0.07
2	2	1	0.13
2	2	2	0.08
3	1	1	0.06
3	1	2	0.08
3	2	1	0.04
3	2	2	0.11

1. Bereken $P(Z=2, X=3|Y=1)$
2. Bereken $P(X=2, Z=1)$
3. Laat U een discrete stochast zijn met mogelijke uitkomsten $u_1, \ldots, u_n$ en laat V een discrete stochast zijn met mogelijke uitkomsten $v_1, \ldots, v_m$. De gezamenlijke kansfunctie is bekend, d.w.z. de kansen $P(U=u, V=v)$ zijn bekend. Wat is de kans P(U=u) uitgedrukt in de kansen $P(U=u, V=v)$?

3 Continue Stochasten

1. Wat is een continue stochast?
2. Gegeven de kansdichtheidsfunctie f_X van een stochast X. Wat is de kans $P(a \leq X \leq b)$ uitgedrukt in de kansdichtheidsfunctie.
3. Wat is de verdelingsfunctie F_X van een stochast X en wat is de relatie met de kansdichtheidfunctie?
4. Van een continue stochast X is de kansdichtheidsfunctie gegeven door:

$$f_X(x) = \begin{cases} Ax^2 & \text{als } 0 \leq x \leq 1 \\ 0 & \text{elders} \end{cases}$$

   waarin A een constante is. Welke waarde *moet* A hebben.

5. Van een continue stochast X is de verwachting is gegeven door:

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

   Bereken de verwachting van de stochast X met de kansdichtheidverdeling zoals gegeven in de vorige vraag.

4 Naive Bayes Classifier (and better)

Om handgeschreven cijfers te herkennen wordt het beeld van een cijfer eerst binair gemaakt (pixelwaarden alleen 0 of 1). De beelden zijn 30 bij 20 pixels. Elk pixel kan gezien worden als een booleaanse stochast X_i voor $i = 1, \ldots, 600$. De klassen zijn $C = 0, .., 9$.

1. Waarom is het ondoenlijk om de gezamenlijke kans functie $P(X_1 = x_1, X_2 = x_2, \ldots, X_{600} = x_{600}, C = c)$ te leren uit voorbeelden?
2. Beschrijf de Naive Bayes Classifier voor dit classificatie probleem. Welke aanname maakt u om de gezamenlijke kansfunctie berekenbaar te maken en op welke manier schat u deze uit de leervoorbeelden?
3. Waarom is de naïeve Bayes veronderstelling voor dit probleem *niet* goed toepasbaar (gebruik je beeldbewerkingskennis).
4. Een SVM is voor deze hoogdimensionale ruimte traag in de praktijk. Daarom is de volgende classifier voorgesteld: verdeel de 600 stochasten in 75 groepen met in iedere groep 8 booleaanse stochasten. Voor de stochasten in een groep wordt verondersteld dat ze afhankelijk zijn, voor de stochasten in 2 verschillende groepen veronderstellen we dat ze onafhankelijk zijn. Hoe bereken je nu de kans $P(X_1 = x_1, X_2 = x_2, \ldots, X_{600} = x_{600}, C = c)$ uit de leervoorbeelden?

5 Multivariate (Vector) Stochasten

1. Geef de definitie van de covariantiematrix $\text{Cov}(\mathbf{X})$ van een vectoriële stochast $\mathbf{X} = (X_1 \cdots X_n)^T$. Wat zijn de afmetingen van deze matrix?
2. Bewijs dat $\text{Cov}(A\mathbf{X}) = A\,\text{Cov}(\mathbf{X})A^T$, daarbij mag u gebruiken dat $E(A\mathbf{X}) = AE(\mathbf{X})$.

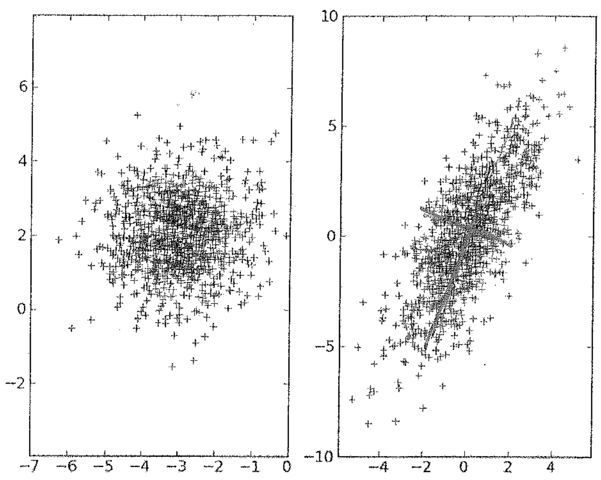

Figure 1: Scatterplot van 1000 realisaties (trekkingen) van $\mathbf{X}$ (links) en van $\mathbf{Y}$ (rechts).

De kansdichtheidsfunctie van een multivariate normaal verdeelde stochast $\mathbf{X} \sim \text{Normal}(\mu, \Sigma)$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$$

1. Wat is Σ en wat is μ en wat zijn hun afmetingen?
2. In figuur 1 zijn de scatterplots van 1000 trekkingen uit 2 normale verdelingen gegeven $X \sim N(\mu_X, \Sigma_X)$ en $Y \sim N(\mu_Y, \Sigma_Y)$. Wat zijn de verwachtingen van beide verdelingen (dwz wat zijn μ_X en μ_Y) ongeveer (immers u moet het niet alleen aflezen uit de figuren maar het blijft probabilistisch natuurlijk...).
3. Een covariantie matrix Σ kan geschreven worden als $U\,D\,U^T$ waarbij U een orthogonale matrix is. Geef de betekenis van deze ontbinding (factorisering) van de covariantiematrix (d.w.z. wat zijn U en D).
4. Leid ook beide covariantie matrices af uit beide scatterplots. Hint 1: de ontbinding uit bovenstaande vraag kan goed van pas komen. Hint 2: herinnert u zich nog uit het eerste deel dat 99.7 % van alle trekkingen zich op een afstand kleiner dan 3σ van het gemiddelde afliggen (voor een scalaire normale verdeling).

6 Principal Component Analysis

1. In figuur 2 een scatterplot 1000 trekkingen uit een 2 dimensionale stochast $\mathbf{X}$. Gegeven is dat het gemiddelde is: $\bar{\mathbf{x}} = \begin{pmatrix}3 & 2\end{pmatrix}^T$ en de schatter van de covariantiematrix:

$$S = \begin{pmatrix}1.88 & 3.56\\ 3.56 & 7.22\end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix}1 & -2\\ 2 & 1\end{pmatrix}\begin{pmatrix}9 & 0\\ 0 & 0.1\end{pmatrix}\frac{1}{\sqrt{5}}\begin{pmatrix}1 & -2\\ 2 & 1\end{pmatrix}^T$$

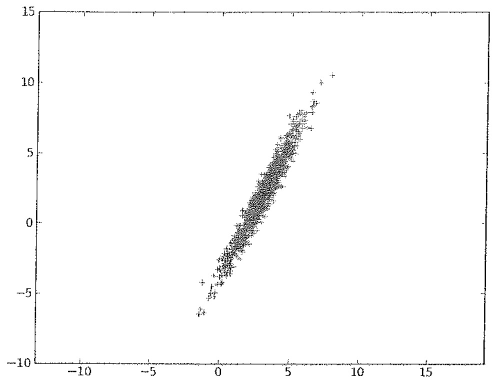

Figure 2: Scatterplot van 1000 realisaties (trekkingen) van **X**.

Als we nu met een PCA de dimensionaliteit van 2 terugbrengen tot 1 kunnen we een willekeurige vector $\mathbf{x} = (a\ b)^T$ als een scalar schrijven. Hoe bereken je die scalar? En hoe benader je $\mathbf{x}$ gegeven alleen de net berekende scalar.

2. Geef in (een schets) van bovenstaande figuur aan wat de fout maximaal kan zijn als je met PCA de data slechts met 1 coördinaat beschrijft.

7 Support Vector Machines

1. In zijn meest eenvoudige vorm tracht een support vector machine een classifier te maken die onderscheid kan maken tussen twee klassen. Als $\mathbf{x}$ een onbekende feature vector is dan wordt deze op basis van een simpele rekenregel toegekend aan klasse A of klasse B:

$$\mathbf{w}^T\mathbf{x} + b \geq 0 \Longrightarrow \text{assign } \mathbf{x} \text{ to class A else to class B}$$

   Wat voor classifier is dit (vernoemd naar de vorm van het scheidingsoppervlak tussen de beide klassen) en wat is de betekenis van de $\mathbf{w}$-vector en de scalar b in bovenstaande rekenregel?

2. Voor een leerset van samples uit zowel klasse A als klasse B zijn er veelal meerdere mogelijkheden om een classifier als hierboven beschreven te maken die de voorbeelden in de leerset juist classificeert. Welke van deze vele mogelijkheden wordt door een SVM uitgerekend?

3. De meest simpele vorm van een SVM zoals hierboven beschreven werkt in veel praktijkgevallen niet. Dan is de vorm van het scheidingsoppervlak niet 'flexibel genoeg'. Op welke wijze maakt een SVM het toch mogelijk om de klassen te scheiden?

4. Op welke manier(en) kan je SVM's inzetten om een meer klassen probleem aan te pakken?

8 Classificatie

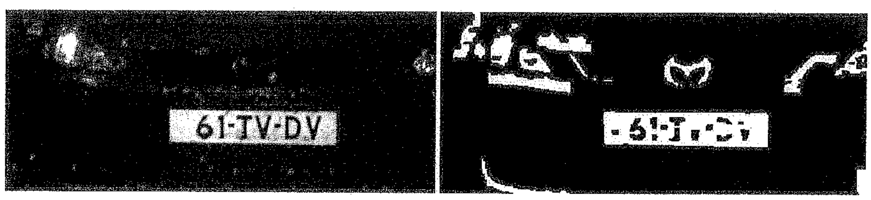

Figure 3: Kentekenplaat detectie.

Beschouw een computer vision toepassing waarin kentekenplaten moeten worden herkend. In Fig.3 is links een infrarood opname te zien van de voorkant van een auto. Omdat kentekenplaten infrarood licht erg goed reflecteren zijn de platen goed zichtbaar.

In een embedded computer vision systeem is het zaak om zo snel mogelijk het hele beeld te reduceren tot maar een beperkt aantal plekken waar de plaat misschien kan zitten. Het werkelijk zoeken en herkennen van letters en cijfers is te complex en te tijdrovend om op het hele beeld toe te passen.

Met een eenvoudige techniek zijn we in staat om het beeld links te reduceren tot een beeld met slechts 2 mogelijke pixelwaarden (zwart of wit) zoals in rechterplaatje te zien valt. Van elke witte 'vlek' (witte pixels die 'aan elkaar vastzitten' en gescheiden zijn van andere witte vlekken) kunnen we zeer snel een aantal kenmerken berekenen:

kenmerk	definitie
X_0	aantal pixels in witte vlek
X_1	aantal pixels in witte vlek + aantal van alle omsloten zwarte pixels (de gaten)
X_2	aantal pixels in kleinste rechthoek die de witte vlek bevat (hoeft niet horizontaal of verticaal te zijn)
X_3	X_1/X_2

Van een klein aantal beelden is een leerset gemaakt van de bovenstaande multivariate stochast. In figuur 4 zijn de scatterdiagrammen van deze 4 dimensionale leerset weergegeven. Een bovenschrift "i-j" geeft aan dat langs de horizontale as X_i staat en langs de vertical as X_j.

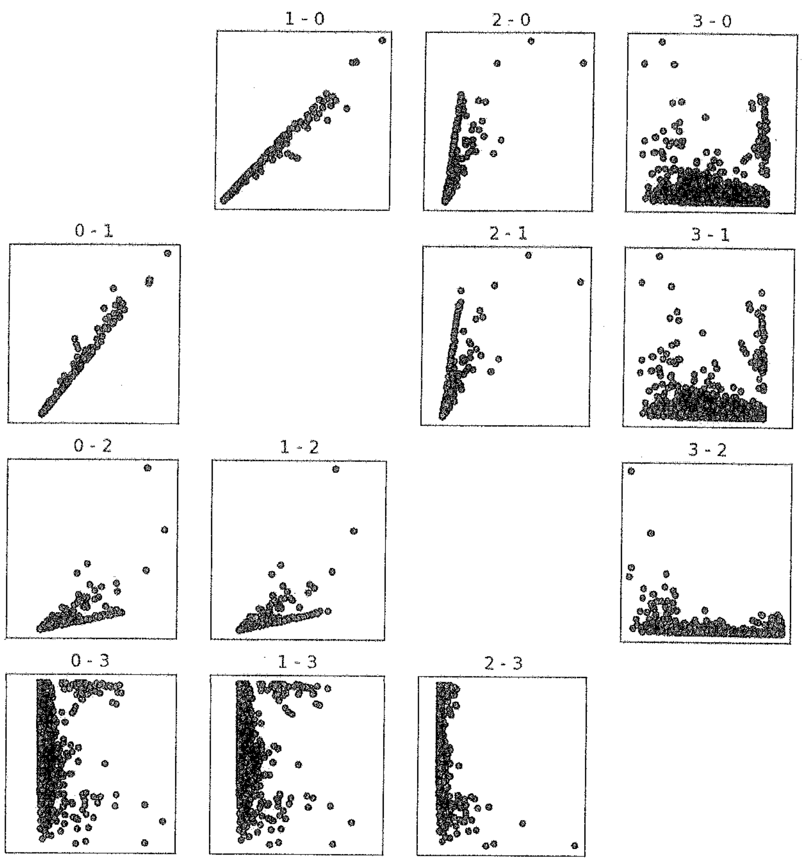

Figure 4: Scatterplots van kenteken vlekken (klasse 1) en niet kentekenvlekken (klasse 2).

Een SVM geleerd op deze dataset haalt inderdaad een score van boven de 99%. Toch wordt die in de praktijk niet gebruikt vanwege de relatieve traagheid van de SVM classifier.

1. Laat aan de hand van de scatterplots zien dat het niet mogelijk is om kentekenplaten van andere 'vlekken' te onderscheiden op basis van slechts 1 van de features.

2. Op basis van twee features kan het wel. Welke features zou u kiezen? En wat voor eenvoudige classifier zou u gebruiken? (dit is een prima moment om zelf een heel eenvoudige classifier visueel te ontwerpen...).

3. Het gaat nooit lukken om op basis van deze features een perfecte classificatie te krijgen. Wat wil dat zeggen voor de *confusion matrix*? Wat is een confusion matrix en hoe denkt u dat hij er voor uw classifier ongeveer uit zal zien?

4. In deze toepassing is het belangrijk dat kentekenplaat vlekken zoveel mogelijk ook als kenteken geclassificeerd worden. Dat daarbij ook wat vaker niet kenteken vlekken als kenteken gezien worden wordt dan voor lief genomen (dat moet dan bij daadwerkelijke herkenning van karakters wel duidelijk worden). Wat voor invloed heeft dit op de gewenste confusion matrix en hoe kunt u dat voor elkaar krijgen met uw classifier uit onderdeel 2.

9 A Game of Jeopardy

Ik geef het antwoord, u formuleert de vraag

1. 42 (just kidding...)
2. $P(A|B) = \frac{P(A)}{P(B)} P(B|A)$
3. $P(AB) = P(A)P(B)$
4. $\forall x, \forall y : P(X = x, Y = y) = P(X = x)P(Y = y)$
5. In dat geval heb je de 'soft-margin' SVM nodig en niet de 'hard-margin' SVM.