Tentamen: Statistisch Redeneren

docent: Rein van den Boomgaard

28 maart 2018

- Maak de opgaven in volgorde op uw antwoordvel
- Vermeld duidelijk naam en studienummer op elk antwoordvel
- Succes !!

1 Kansrekening

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 gezamelijke kansdichtheid (joint distribution) is $p_{XYZ}(x,y,z) = P(X=x, Y=y, Z=z)$ en is geënumereerd in onderstaande tabel:

x	y	z	$p_{XYZ}(x,y,z)$
1	1	1	0.11
1	1	2	0.09
1	2	1	0.08
1	2	2	0.04
2	1	1	0.13
2	1	2	0.07
2	2	1	0.15
2	2	2	0.08
3	1	1	0.06
3	1	2	0.12
3	2	1	0.04
3	2	2	0.03

1. Bereken $P(Y=1)$ en $P(X=3, Z=2)$
2. Bereken $P(X=2|Y=1)$.

2 Continue Stochasten

1. De kansdichtheid functie voor continue stochast X is f_X. Hoe bereken je de cumulatieve distributiefunctie F_X uit f_X.
2. Geef de kans $P(a \leq X \leq b)$ in termen van de distributiefunctie F_X.
3. Een verdeling wordt bepaald door zijn kansdichtheidsfunctie:

$$f_X(x) = \begin{cases} A \sin x & : 0 \leq x \leq \pi \\ 0 & : \text{elsewhere} \end{cases}$$

   Laat met een berekening zien dat de waarde van de constante A gelijk $1/2$ is.
4. Bereken de verwachting voor de stochast X met kansdichtheidsverdeling zoals hierboven gegeven. Hint: bereken de afgeleide van $\sin x - x \cos x$.

3 Naive Bayes Classifier

Stel we meten n features $X_1, X_2, \ldots, X_n$ van objecten die we willen classificeren als $C = 1$, $C = 2$ of $C = 3$. Neem aan dat alle features booleaanse stochasten zijn (d.w.z. $X_i = 0$ of $X_i = 1$).

1. We kennen aan een onbekend object met kenmerken $x_1, \ldots, x_n$ de klasse k toe waarvoor $P(C = k | X_1 = x_1, \ldots, X_n = x_n)$ maximaal is. Herschrijf deze term met de Bayesregel.
2. Geef de formule voor de Naïve Bayes Classifier voor het bepalen van de klasse gegeven de kenmerk waarden $x_1, \ldots, x_n$.
3. Welke kansen moeten geschat worden uit de leervoorbeelden om deze classifier te kunnen gebruiken? En hoe schat u die kansen uit leervoorbeelden?

4 Multivariate Stochasten

1. Geef de definitie van de covariantie matrix $\mathrm{Cov}(\mathbf{X})$ van de vectoriele stochast $\mathbf{X}$ in termen van de covariantie van de scalaire stochasten. Wat zijn de afmetingen van deze matrix?
2. Bewijs dat $\mathrm{Cov}(A\mathbf{X}) = A\,\mathrm{Cov}(\mathbf{X})A^\intercal$. Hint: vermijd het uitschrijven van de elementen van de matrix maar gebruik de vectoriele definitie van de covariantie matrix.
3. Een mooie eigenschap van de normale verdeling is dat elke lineaire combinatie van normaal verdeelde stochasten ook normaal verdeeld is. Maak duidelijk dat als $\mathbf{X}$ een multivariate normale verdeling is, dat dan dus ook $A\mathbf{X}$ een multivariate normale verdeling is.
4. We kunnen eenvoudig trekkingen uit de normale verdeling met verwachting 0 en covariantie matrix I (eenheidsmatrix) maken. Een stochast met deze verdeling noemen we $\mathbf{X}$. We willen nu een stochast $\mathbf{Y} = A\mathbf{X}$ 'maken' met covariantie matrix $\Sigma = UDU^\intercal$ (eigendecompositie). Welke matrix A moet ik kiezen, uitgdrukt in U en D?

5 Principal Component Analysis

1. Leg uit in drie zinnen hoe je met PCA in staat bent om in veel praktijk situaties de dimensionaliteit van een probleem te verlagen.
2. Gegeven is een stochast $\mathbf{X}$ met verwachting $\mathrm{m} = E(\mathbf{X}) = (3 \quad 2)^\intercal$ en covariantiematrix:

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

   Als we nu met een PCA de dimensionaliteit van 2 terugbrengen tot 1 kunnen we een willekeurige vector $\mathbf{x} = (a\ b)^\intercal$ met één scalar q representeren. Hoe bereken je die scalar?
3. Hoe benader je $\mathbf{x}$ gegeven alleen de net berekende scalar q. Noem je benadering $\hat{\mathbf{x}}$.
4. Wat is ongeveer de grootte van de fout $\|\mathbf{x} - \hat{\mathbf{x}}\|$?
5. (Alleen als je veel tijd over hebt... voor bonus). Bereken $E(\|\mathbf{X} - \hat{\mathbf{X}})\|^2)$ en $\mathrm{Var}(\|\mathbf{X} - \hat{\mathbf{X}})\|^2)$ waarin $\hat{\mathbf{X}}$ de benadering is van stochast $\mathbf{X}$. Alleen in deze opgave mag u er voor de eenvoud vanuit gaan dat $E(\mathbf{X}) = 0$.

6 Linear Regression

Gegeven de training set $(x^{(j)}, y^{(j)})$ voor $j = 1, \ldots, m$ zoals geschetst in figuur.

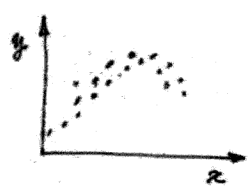

Figure 1: Linear Regression

1. Schets (bij benadering) de functie $h_\theta(x) = \theta_0 + \theta_1 x$ voor de optimale waarden van θ_0 and θ_1 die gevonden worden door minimalisatie van de kosten function.
2. Wat is het probleem met deze regressie hypothese? Is het een 'overfitting' probleem (variantie probleem) of een 'underfitting' probleem (bias probleem)? Kun je dit probleem detecteren in de leer fase? Of pas in de test fase. Waarom?
3. Beschouw de volgende hyposthese:

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

Schets ook deze hypothese functie nadat de parameters zijn gevonden door minimalisatie van de kost functie.

4. Iemand suggereert weer een andere hypothese functie:

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 + \theta_5 x^5 + \theta_6 x^6 + \theta_7 x^7$$

Wat zal het probleem zijn als we de parameters in dit model leren aan de hand van de gegeven data? Is dit probleem detecteerbaar in de leerfase of pas in de testfase? En wat is een standaard oplossing van dit probleem (zonder een andere hypothese te kiezen).

7 Logistic Regression

Beschouw de dataset in bijgaand figuur. De twee klassen moeten onderscheiden worden met logistic regression.

Figure 2: Logistic Regression

1. De hypothese functie is:

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

waarbij g de sigmoid functie is. Wat is de waarde van de hypothese functie op de scheidslijn (decision boundary)?

2. Schets (bij benadering) deze scheidslijn nadat de optimale parameters zijn gevonden. Geef ook een vergelijking voor deze scheidslijn.
3. Beschrijf in 3 zinnen hoe die optimale parameters worden gevonden (geleerd).

8 Neurale Netwerken

1. Leg uit waarom het niet mogelijk is om in een netwerk zonder hidden layers (dus alleen een input en output laag) niet mogelijk is om het xor probleem op te lossen.
2. Teken een neuraal netwerk met 2 input nodes, 2 hidden layers met ieder 2 nodes en een output layer met 1 node.
3. Geef ook de dimensies van alle weegfactor matrices Θ.