Tentamen Statistisch Redeneren

Rein van den Boomgaard

28 Maart 2017

- Graag bondige antwoorden! Een goed antwoord dat ik moet zoeken tussen informatie die niet van belang is wordt fout gerekend.
- Veel 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 gezamenlijke kansdichtheid (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	0.11
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 $P(X=x)$ voor $x = 1,2,3$.
2. Wat is de *definitie* van de verwachting $E(X)$ van X. Bereken $E(X)$ voor de stochast X zoals gedefinieerd in de kanstabel.
3. Wat is de definitie van de voorwaardelijke kans $P(A|B)$?
4. Bereken $P(Y=2|X=3, Z=1)$.

2 Continue Stochasten

1. Gegeven de stochast X met kansdichtheidsfunctie f_X. Wat is de kans $P(a \leq X \leq b)$ uitgedrukt in de kansdichtheidsfunctie?
2. Wat is de definitie van de verwachting van een continue stochast?
3. We beschouwen nu n identiek verdeelde en onafhankelijke stochasten $X_1, \ldots, X_n$. Wat is de verwachting van de som $S = X_1 + \cdots + X_n$ en wat is de verwachting van het gemiddelde $M = (X_1 + \cdots + X_n)/n$?

4. Voor dezelfde stochasten als in de vraag hierboven: wat is de variantie van de som S en wat is de variantie van het gemiddelde M?

In enkele van de vragen hierna wordt verwezen naar een artikel uit het wetenschappelijk journal PAMI. Lees dit artikel pas nadat je in de vraag een verwijzing tegenkomt!!

2.2.2 Bayesian Classifier with the Histogram Technique

The Bayesian decision rule for minimum cost is a well-established technique in statistical pattern classification [17]. Using this decision rule, a color pixel $\mathbf{x}$ is considered as a skin pixel if

$$\frac{p(\mathbf{x}|skin)}{p(\mathbf{x}|nonskin)} \geq \tau, \tag{1}$$

where $p(\mathbf{x}|skin)$ and $p(\mathbf{x}|nonskin)$ are the respective class-conditional pdfs of skin and nonskin colors and τ is a threshold. The theoretical value of τ that minimizes the total classification cost depends on the a priori probabilities of skin and nonskin and various classification costs; however, in practice τ is often determined empirically. The class-conditional pdfs can be estimated using histogram or parametric density estimation techniques. The Bayesian classifier with the histogram technique has been used for skin detection by Wang and Chang [12] and Jones and Rehg [4].

2.2.3 Gaussian Classifiers

The class-conditional pdf of skin colors is approximated by a parametric functional form, which is usually chosen to be a unimodal Gaussian [2], [14] or a mixture of Gaussians [13], [15]. In the case of the unimodal Gaussian model, the skin class-conditional pdf has the form:

$$\begin{aligned} p(\mathbf{x}|skin) &= g(\mathbf{x}; \mathbf{m}_s, \mathbf{C}_s) \\ &= (2\pi)^{-d/2} |\mathbf{C}_s|^{-1/2} exp\left\{ -\frac{1}{2}(\mathbf{x}-\mathbf{m}_s)^T \mathbf{C}_s^{-1} (\mathbf{x}-\mathbf{m}_s) \right\}, \end{aligned} \tag{2}$$

Skin Segmentation Using Color Pixel Classification: Analysis and Comparison

Son Lam Phung, *Member, IEEE*,
Abdesselam Bouzerdoum, *Sr. Member, IEEE*,
and Douglas Chai, *Sr. Member, IEEE*

Abstract—This paper presents a study of three important issues of the color pixel classification approach to skin segmentation: color representation, color quantization, and classification algorithm. Our analysis of several representative

where d is the dimension of the feature vector, $\mathbf{m}_s$ is the and $\mathbf{C}_s$ is the of the skin class. If we assume that the nonskin class is uniformly distributed, the Bayesian rule in (1) reduces to the following: a color pixel $\mathbf{x}$ is considered as a skin pixel if

$$(\mathbf{x}-\mathbf{m}_s)^T \mathbf{C}_s^{-1} (\mathbf{x}-\mathbf{m}_s) \leq \tau, \tag{3}$$

where τ is a threshold and the left hand side is the squared Mahalanobis distance. The resulting decision boundary is an ellipse in 2D space and an ellipsoid in 3D space.

3 Multivariate Stochasten (Multivariate Random Variables)

1. De vector $\mathbf{x}$ in formule (1) in het geciteerde artikel is een waarde van de stochast $\mathbf{X} = (RGB)^\intercal$ welke een kleur beschrijft met de drie kleur waarden R, G en B. Wat is de verwachting $E(\mathbf{X})$ (in woorden en formule).

2. Geef de definitie van de covariantie matrix $\mathrm{Cov}(\mathbf{X})$ van de vectoriele stochast $\mathbf{X}$. Wat zijn de afmetingen van deze matrix?

3. Bewijs dat $E(A\mathbf{X}) = AE(\mathbf{X})$. Daarbij mag u gebruik maken van de eigenschappen van de verwachting van scalaire stochasten ($E(aX + bY) = \cdots$).

4 Bayes Classificatie

1. In sectie 2.2.2 van het artikel wordt de Bayesian classifier gegeven in formule (1):

$$\frac{p(\mathbf{x}|\mathrm{skin})}{p(\mathbf{x}|\mathrm{nonskin})} \geq \tau$$

en wordt beweerd dat τ voor de "minimum error classifier" afhankelijk is van de a priori kansen. Geef de afleiding van bovenstaande classifier en druk τ uit in $P(\mathrm{skin})$ en $P(\mathrm{nonskin})$.

2. Laat zien dat de aanname van een uniforme distributie van de nonskin klasse leidt tot formule (3) voor de classifier.

3. Met welke aanname maak je van een Bayes Classifier een *Naive* Bayes Classifier? Wat is het voordeel van de Naive Bayes Classifier boven de Bayes classifier?

5 Principal Component Analysis

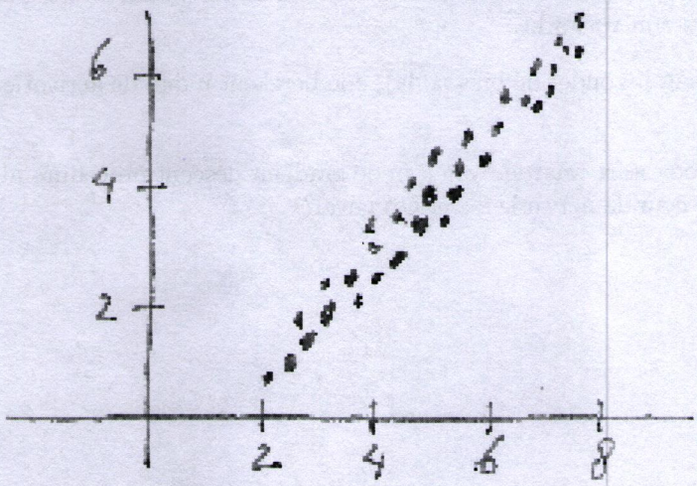
1. De belangrijke twee formules in PCA zijn:

$$\begin{aligned} \mathbf{y}_k &= U_k^\intercal(\mathbf{x} - \mathbf{m}) \\ \mathbf{x} &\approx U_k\mathbf{y}_k + \mathbf{m} \end{aligned}$$

waarin U_k de matrix is bestaande uit de eerste k kolommen van de matrix U. Geef de betekenis van $\mathbf{y}_k$, $\mathbf{x}$, U en $\mathbf{m}$?

2. Waarop baseert u in de praktijk de keuze van k?

3. Voor de dataset gegeven in onderstaande figuur geef een schatting voor U_1 en voor $\mathbf{m}$.



4. Voor $\mathbf{x} = (7\ 5)^\intercal$ bereken $\mathbf{y}_1$ en $U_1\mathbf{y}_1 + \mathbf{m}$.

6 Logistic Regression Classifier

1. De logistic regression classifier is net als de Bayes classifier gebaseerd op maximalisatie van de a posteriori kans

$$P(Y = y|\mathbf{X} = \mathbf{x})$$

over alle mogelijke klassen y. Waarin verschilt de logistic regression classifier essentieel met de Bayes classifier?

2. Welke hypothese functie $h_\theta(\mathbf{x})$ wordt gebruikt in logistic regression?

3. De kosten functie voor logistic regression is:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta(\mathbf{x}^{(i)}) + (1 - y^{(i)})(1 - h_\theta(\mathbf{x}^{(i)}))$$

Waarom wordt bovenstaande kosten functie gebruikt en niet de ietwat eenvoudigere gemiddelde kwadratische fout functie?

4. Welke regularisatie term moet u aan de kosten functie toevoegen om de θ waarden zo klein mogelijk te houden?

5. Waarvoor dient regularisatie? Hoe bepaalt u in de praktijk de mate van regularisatie (d.w.z. de waarde van de parameter λ die u in de regularisatie term in voorgaande vraag ongetwijfeld ingevoerd heeft).

7 Neurale Netwerken

1. Schets een 4 laags feed forward neuraal netwerk met 2, 4, 3, en 1 node in de lagen (gegeven van input laag tot output laag). Schets ook de bias inputs en de connecties tussen de nodes!

2. Wat zijn de afmetingen van alle θ matrices in het netwerk? Daarbij verondersteld dat de weegfactoren voor de bias nodes in de θ matrices zijn verwerkt.

3. Als $\mathbf{a}^{(l)}$ de activatie vector is van laag l (zonder de bias node), hoe berekent u dan de activatie $\mathbf{a}^{(l+1)}$ van laag $l+1$.

4. Bij logistic regression is de keuze voor start waarden van θ in de gradient descent procedure niet van groot belang. Waarom is dat voor neurale netwerken wel het geval?