Kompletteringar till föreläsning i Täthetsskattning. I

2nd November 2005

1 ML-skattning av väntevärdesvektor i en normalfördelning

Vi antar att vi har ett antal $\mathbf{x}$ som är dragna oberoende från varandra från en multivariat normalfördelning med okänd väntevärdesvektor, $\mathbf{m}$, och inverterbar (och känd) kovariansmatris $\mathbf{C}$. Vi ska här se hur vi ska ta fram en bra skattning av $\mathbf{m}$.

Låt $\mathcal{X}$ beteckna vår mängd med N st. exempel, dvs $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$. Om $\mathbf{x}_n$ är normalfördelad $\mathbf{x} \sim N(\mathbf{m}, \mathbf{C})$ så har den täthetsfunktionen

$$p(\mathbf{x}_n|\mathbf{m},\mathbf{C}) = \frac{1}{(2\pi)^{d/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{x}_n - \mathbf{m}\right)^T \mathbf{C}^{-1}\left(\mathbf{x}_n - \mathbf{m}\right)\right) \tag{1}$$

Per definition ges ML-skattningen av

$$\hat{\mathbf{m}}_{ML} = \arg\max_{\mathbf{m}} p(\mathcal{X}|\mathbf{m},\mathbf{C}) = \arg\max_{\mathbf{m}} p(\mathbf{x}_1, ..., \mathbf{x}_N|\mathbf{m},\mathbf{C}) = \arg\max_{\mathbf{m}} \prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{m},\mathbf{C}) \tag{2}$$

Den sista likheten fås pga av att exemplen är oberoende. Observera att vi ser $p(\mathcal{X}|\mathbf{m},\mathbf{C}) = p(\mathbf{x}_1, ..., \mathbf{x}_N|\mathbf{m},\mathbf{C})$ som en funktion av $\mathbf{m}$. Vektorerna $\mathbf{x}_1, ..., \mathbf{x}_N$ är däremot fixa.

Då vi endast är intresserade av vilken vektor $\mathbf{m}$ som maximerar $p(\mathcal{X}|\mathbf{m},\mathbf{C})$ så kan vi lika gärna söka den vektor som minimerar negativa logaritmen samt bortse från eventuella termer som inte beror av $\mathbf{m}$. Efter förenkling av $-\ln p(\mathbf{x}_n|\mathbf{m},\mathbf{C})$ får vi

$$\hat{\mathbf{m}}_{ML} = \arg\min_{\mathbf{m}} \underbrace{\frac{1}{2}\sum_{n=1}^{N}\left(\mathbf{x}_n - \mathbf{m}\right)^T \mathbf{C}^{-1}\left(\mathbf{x}_n - \mathbf{m}\right)}_{J(\mathbf{m})} \tag{3}$$

För minimat till $J(\mathbf{m})$ gäller att $\frac{\partial J}{\partial \mathbf{m}} = 0$. Vi utvecklar först J som

$$\begin{aligned} J &= \frac{1}{2}\sum_{n=1}^{N}\left(\mathbf{x}_n - \mathbf{m}\right)^T \mathbf{C}^{-1}\left(\mathbf{x}_n - \mathbf{m}\right) = \frac{1}{2}\sum_{n=1}^{N}\mathbf{x}_n^T\mathbf{C}^{-1}\mathbf{x}_n - \sum_{n=1}^{N}\mathbf{m}^T\mathbf{C}^{-1}\mathbf{x}_n + \frac{1}{2}\sum_{n=1}^{N}\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m} = \\ &= \frac{1}{2}\sum_{n=1}^{N}\mathbf{x}_n^T\mathbf{C}^{-1}\mathbf{x}_n - \mathbf{m}^T\left(\sum_{n=1}^{N}\mathbf{C}^{-1}\mathbf{x}_n\right) + \frac{N}{2}\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m} \end{aligned} \tag{4}$$

Gradienten finner vi sen enkelt genom att utnyttja våra deriveringsregler för vektoruttryck: (Observera att vi kan skriva, mest som en hjälp för minnet, tex $\mathbf{x}_n^T\mathbf{C}^{-1}\mathbf{m} = tr(\mathbf{x}_n^T\mathbf{C}^{-1}\mathbf{m})$.)

$$\frac{\partial J}{\partial \mathbf{m}} = 0 - \left(\sum_{n=1}^{N} \mathbf{C}^{-1}\mathbf{x}_n\right) + \frac{N}{2}(\mathbf{C}^{-1} + \mathbf{C}^{-T})\mathbf{m} = -\left(\sum_{n=1}^{N} \mathbf{C}^{-1}\mathbf{x}_n\right) + N\mathbf{C}^{-1}\mathbf{m} \qquad (5)$$

där $\mathbf{C}^{-T}$ betyder transponatet av inversen. Pga att $\mathbf{C}$ är symmetrisk är även $\mathbf{C}^{-1}$ symmetrisk och därför gäller alltså att $\mathbf{C}^{-T} = \mathbf{C}^{-1}$.

Löser vi $\frac{\partial J}{\partial \mathbf{m}} = 0$ så får vi till slut

$$\hat{\mathbf{m}}_{ML} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n, \qquad (6)$$

dvs det gamla hederliga medelvärdet. Notera f.ö. att lösningen inte beror av $\mathbf{C}$.

2 MAP-skattning

Om vi har en a priori-fördelning för $\mathbf{m}$ som är Gaussisk med väntevärde $\mathbf{m}_0$ och kovariansmatris $\mathbf{C}_m$, dvs

$$p(\mathbf{m}) = \frac{1}{(2\pi)^{-d/2}|\mathbf{C}_m|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{m} - \mathbf{m}_0\right)^T \mathbf{C}_m^{-1}\left(\mathbf{m} - \mathbf{m}_0\right)\right) \qquad (7)$$

så går det efter en stunds räknande att visa att MAP-skattningen av $\mathbf{m}$ ges av

$$\hat{\mathbf{m}}_{MAP} = \left(N\mathbf{C}^{-1} + \mathbf{C}_m^{-1}\right)^{-1}\left(\mathbf{C}^{-1}\sum_{n=1}^{N}\mathbf{x}_n + \mathbf{C}_0\mathbf{m}_0\right) \qquad (8)$$

Vi kan se flera samband med ML-skattningen:

- Låt $N \to \infty$, dvs antalet exempel växer mot oändligheten. Vi kan skriva om MAP-skattningen ovan som

$$\begin{aligned}\hat{\mathbf{m}}_{MAP} &= \left(\mathbf{C}^{-1} + \frac{1}{N}\mathbf{C}_m^{-1}\right)^{-1}\left(\mathbf{C}^{-1}\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n + \frac{1}{N}\mathbf{C}_0\mathbf{m}_0\right) \\ &\to \left(\mathbf{C}^{-1}\right)^{-1}\left(\mathbf{C}^{-1}\hat{\mathbf{m}}_{ML}\right) = \hat{\mathbf{m}}_{ML}\end{aligned} \qquad (9)$$

  Alltså, MAP- och ML-skattningarna är asymptotiskt identiska då antalet exempel växer mot oändligheten.

- Om $\mathbf{C}_m$ är "stor"[1], alternativt, $\mathbf{C}$ är "liten", så kan MAP- och ML-skattningarna vara mycket nära varandra. Tex, sätt $\mathbf{C}_m = \sigma_m^2\mathbf{I}$, där $\mathbf{I}$ är enhetsmatrisen och σ_m^2 är ett

[1] I detta sammanhang betyder detta att matrisens egenvärden alla är stora

stort tal. Vi får då återigen

$$
\begin{aligned}
\hat{\mathbf{m}}_{MAP} &= \left(\mathbf{C}^{-1} + \frac{1}{N}\mathbf{C}_m^{-1}\right)^{-1} \left(\mathbf{C}^{-1}\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n + \frac{1}{N}\mathbf{C}_0\mathbf{m}_0\right) \\
&\rightarrow \left(\mathbf{C}^{-1}\right)^{-1} \left(\mathbf{C}^{-1}\frac{1}{N}\hat{\mathbf{m}}_{ML}\right) = \hat{\mathbf{m}}_{ML}
\end{aligned}
\tag{10}
$$