

DS-GA 1018: Homework 3

Due Monday October 28th at 5:00 pm

Problem 1 (19 points): Consider the latent space model we presented in class defined by:

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{w}_t \quad (1)$$

$$\mathbf{x}_t = \mathbf{C}\mathbf{z}_t + \mathbf{v}_t \quad (2)$$

where the latent space \mathbf{z} has dimension d and the data \mathbf{x} has dimension n . Our noise is being drawn from $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q})$ and $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R})$. Assume that $n = d$ and that we have $\mathbf{A} = \mathbb{I}$, $\mathbf{C} = \mathbb{I}$, and $\mathbf{Q} = 0 \cdot \mathbb{I}$. Finally, assume that $\Sigma_0 \rightarrow \infty$. We make no specific assumption for \mathbf{R} or μ_0 . We want to show that the smoothed mean $\mu_{t|T}$ simplifies to the mean of the data $\frac{1}{T} \sum_{i=1}^T \mathbf{x}_i$. This problem will walk you through the steps:

- i. **(5 points):** Show that $\mu_{1|1} = \mathbf{x}_1$ and $\Sigma_{1|1} = \mathbf{R}$. You will need to take the limit $\Sigma_0 \rightarrow \infty$. *Hint: it will be useful to know that $(\mathbb{I} + \mathbf{M})^{-1} = \sum_{i=0}^{\infty} (-1)^i \mathbf{M}^i$, where \mathbf{M} is an arbitrary matrix.*
- ii. **(2 points):** Explain the intuition behind the result for $\mu_{1|1}$ and $\Sigma_{1|1}$.
- iii. **(4 points):** Assume that $\mu_{t-1|t-1} = \frac{1}{t-1} \sum_{i=1}^{t-1} \mathbf{x}_i$ and that $\Sigma_{t-1|t-1} = \frac{1}{t-1} \mathbf{R}$ (this is consistent with your previous result). Show that $\mu_{t|t} = \frac{1}{t} \sum_{i=1}^t \mathbf{x}_i$ and that $\Sigma_{t|t} = \frac{1}{t} \mathbf{R}$.
- iv. **(6 points):** Finally, show that $\mu_{t|T} = \frac{1}{T} \sum_{i=0}^T \mathbf{x}_i$ and $\Sigma_{t|T} = \frac{1}{T} \mathbf{R}$ for all t .
- v. **(2 points):** Explain the intuition behind the final result.

Problem 2 (20 points): In lecture we discussed the importance of the initial parameter values for the results of the expectation-maximization algorithm. In this problem we will quantitatively demonstrate that importance. Consider the EM algorithm for an LDS system like the one in class. Imagine that we have an initial guess for our parameters:

$$\boldsymbol{\theta}_0 = \{\boldsymbol{\mu}_{0,0}, \boldsymbol{\Sigma}_{0,0}, \mathbf{A}_0, \mathbf{Q}_0, \mathbf{C}_0, \mathbf{R}_0\}. \quad (3)$$

We have also set our distribution $q_1(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_0)$. In other words, q_1 is the distributions set by the Kalman smoothing means and covariances (the standard choice).

- i. **(3 points):** Consider the case where $\mathbf{A}_0 = 0 \cdot \mathbb{I}$ and $\mathbf{C}_0 = 0 \cdot \mathbb{I}$. what will the value of $\boldsymbol{\mu}_{0,1}$ be in terms of the data in \mathbf{X} and the parameters in $\boldsymbol{\theta}_0$? *For all the subproblems, don't forget to consider how our choice of $\boldsymbol{\theta}_0$ affects terms like $\boldsymbol{\mu}_{t|t-1}$.*
- ii. **(3 points):** Still considering the case where $\mathbf{A}_0 = 0 \cdot \mathbb{I}$ and $\mathbf{C}_0 = 0 \cdot \mathbb{I}$, what will the value of $\boldsymbol{\Sigma}_{0,1}$ be in terms of the data in \mathbf{X} and the parameters in $\boldsymbol{\theta}_0$?
- iii. **(4 points):** Still considering the case where $\mathbf{A}_0 = 0 \cdot \mathbb{I}$ and $\mathbf{C}_0 = 0 \cdot \mathbb{I}$, what will the value of \mathbf{A}_1 be in terms of the data in \mathbf{X} and the parameters in $\boldsymbol{\theta}_0$?
- iv. **(4 points):** Still considering the case where $\mathbf{A}_0 = 0 \cdot \mathbb{I}$ and $\mathbf{C}_0 = 0 \cdot \mathbb{I}$, what will the value of \mathbf{Q}_1 be in terms of the data in \mathbf{X} and the parameters in $\boldsymbol{\theta}_0$?
- v. **(4 points):** Still considering the case where $\mathbf{A}_0 = 0 \cdot \mathbb{I}$ and $\mathbf{C}_0 = 0 \cdot \mathbb{I}$, what will the value of \mathbf{C}_1 be in terms of the data in \mathbf{X} and the parameters in $\boldsymbol{\theta}_0$?
- vi. **(2 points):** Given these results, what is the solution to $\boldsymbol{\mu}_{0,n}$, $\boldsymbol{\Sigma}_{0,n}$, \mathbf{A}_n , \mathbf{C}_n , and \mathbf{Q}_n ?