Data Engineering 424: Practical week 9

Name: Dgibrilly M. Mutabazi   SU: 23765518

Declaration

I hereby declare that this work is my own. No assistance has been received from any persons, demi, lecturer or student. Assistance was however received from ChatGPT. It was used for debugging and improved code quality and improved vocabulary where needed.

Derivations and Theoretical Insights   -(Refers to: Q3(b–c), Q4(b–c) of Week 9 + Q4(c) of Week 8)

Week 9: Question (3b) and (3c)

In this section, I present an overview of the derivations leading to the MAP estimate in Question 3(c), based on the prior and likelihood terms described in Question 3(b).

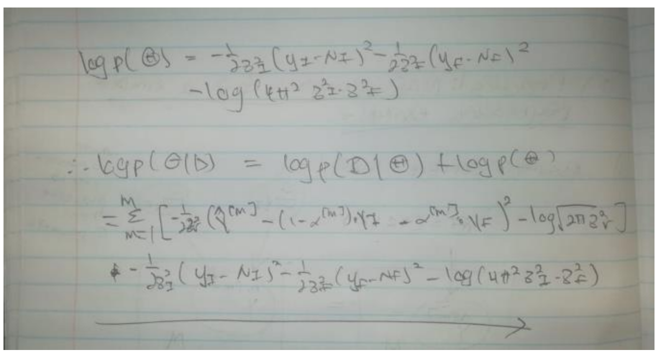


figure 1

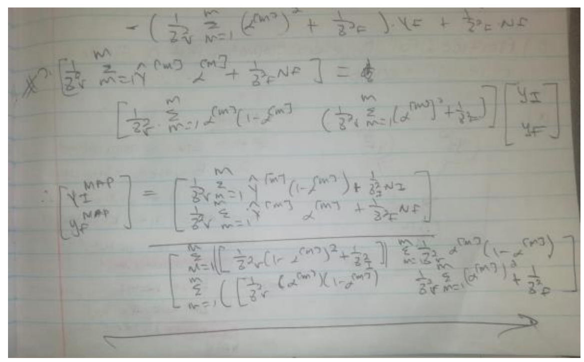
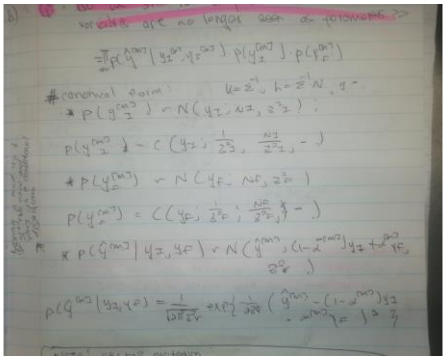


figure 2

The goal of MAP estimation is to find the parameters which maximise the distribution over the parameters given the data. This result written as p(θ | D) can be written as the product of the likelihood of the data (p(D| θ)) with the prior (p(θ)) of the parameters. The log (base e) of the posterior is taken and results in the summation of the log of the posterior distribution with the log of the prior. The final result after applying the log transformation and expanding the quantities in terms of the the Gaussian denisity is shown in the figure 1 above. Using this result the derivative is taken with respect to parameters yI and yF and after manipulating the results such that the knowns and unknowns are grouped together, the result obtained for the MAP estimates of yI and yF are shown in figure 2.

Week 9: Question (4b) and (4c)

Similarly, the full Bayesian posterior derivation in Question 4(c) builds upon the canonical form representations established in Question 4(b).



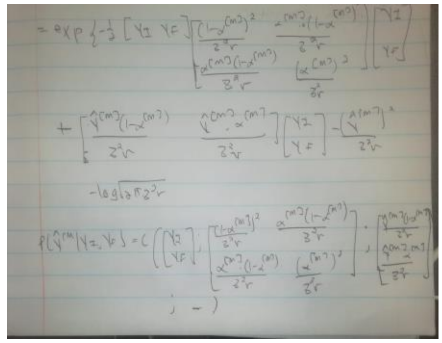


Figure 3 Figure 4

In the full Bayesian framework, I began by modeling the parameters yI and yF as random variables, rather than fixed unknown parameters. This approach treats uncertainty about the line's endpoints probabilistically. Specifically, I assigned independent Gaussian priors to both variables. Each prior was then rewritten in canonical (information) form, as shown in figure 3 above. For the likelihood term p(y^[m]|yI,yF) I used the measurement model from Week 8, rewriting this as a conditional Gaussian distribution in mean-covariance form, simplifying it and then converting it to a canonical term as shown in figure 4, by identifying the information matrix K, and information vector h. After this each term factor was in canonical form, then came the final step.

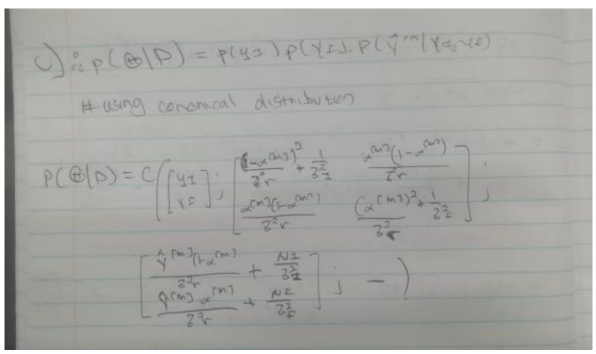


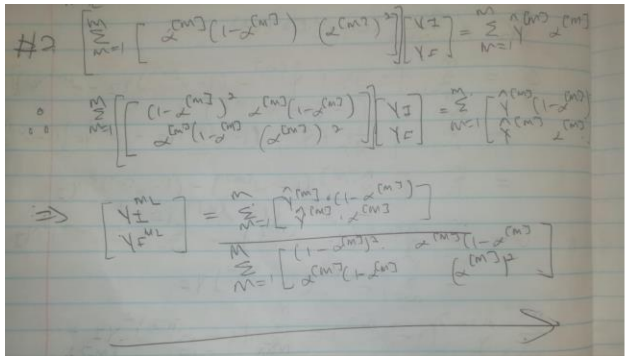
Figure 5

Because each factor was in canonical form, finding the posterior distribution p(θ | D) was found by taking the product of the likelihood and priors that were in canonical form to yield the result in figure 5 above.

Week 9 and 8: Comparison of Question (3c) and (4c) of week 9 with Question (4c) week 8

Similarities and Differences.

The similarities between question (4c) of week 8 (shown in Figure 6 below) and question (3c) of week 8 is that the form of the MAP estimates and ML estimates are similar. However the biggest difference is the addition of the regularization parameters added in the MAP estimates for yI and yF. This parameters is as a result of the prior information that MAP estimates take into account and are responsible to ensure that the estimates of the parameters yI and yF take on reasonable values even if there are few data points. This is a limitation the ML estimates. Below is the ML estimate for the parameters yI and yF.


$$\#2 \quad \left[\sum_{m=1}^{M} \left[\alpha^{[m]}(1-\alpha^{[m]}) \quad (\alpha^{[m]})^2 \right]\right] \begin{bmatrix} y_I \\ y_F \end{bmatrix} = \sum_{m=1}^{M} \hat{y}^{[m]} \alpha^{[m]}$$

$$\therefore \sum_{m=1}^{M} \begin{bmatrix} (1-\alpha^{[m]})^2 & \alpha^{[m]}(1-\alpha^{[m]}) \\ \alpha^{[m]}(1-\alpha^{[m]}) & (\alpha^{[m]})^2 \end{bmatrix} \begin{bmatrix} y_I \\ y_F \end{bmatrix} = \sum_{m=1}^{M} \begin{bmatrix} \hat{y}^{[m]}(1-\alpha^{[m]}) \\ \hat{y}^{[m]}\alpha^{[m]} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} y_I^{ML} \\ y_F^{ML} \end{bmatrix} = \frac{\sum_{m=1}^{M} \begin{bmatrix} \hat{y}^{[m]}(1-\alpha^{[m]}) \\ \hat{y}^{[m]}\alpha^{[m]} \end{bmatrix}}{\sum_{m=1}^{M} \begin{bmatrix} (1-\alpha^{[m]})^2 & \alpha^{[m]}(1-\alpha^{[m]}) \\ \alpha^{[m]}(1-\alpha^{[m]}) & (\alpha^{[m]})^2 \end{bmatrix}}$$

Figure 6 question (4c) week 8

Effect of Number of Measurements - (Refers to: Q3(e), Q4(f) of Week 9 + Q4(f) of Week 8)

To investigate the influence of data availability, I conducted experiments in Question 3(e), Question 4(f) of Week 9, and Question 4(f) of Week 8 by varying the number of measurements M. The objective was to observe how ML, MAP, and full Bayesian estimates behave under different data volumes.

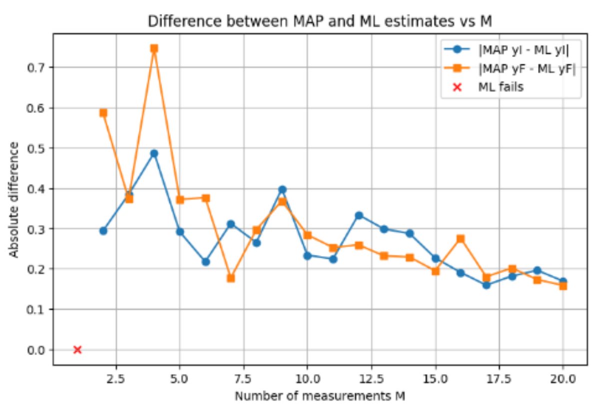


figure 7

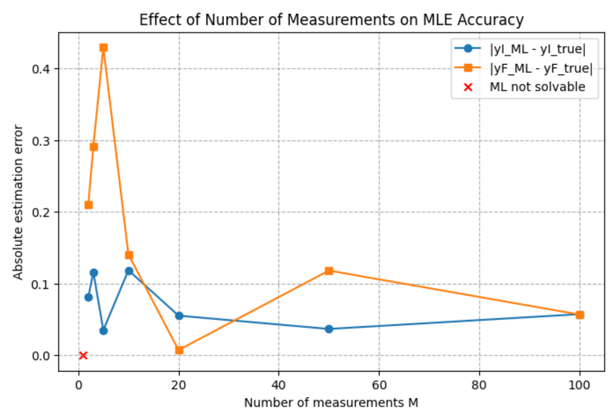


figure 8

Week 9 - Question (3e): MAP Estimation

In my experiments, shown in figure 7, I varie'd the number of measurements M from 1 to 20 (keeping σ_v^2=0.01 and Gaussian priors fixed). I observed that my MAP estimator remains well-defined even when M=1, thanks to the prior, whereas the ML estimator fails at M=1 and only becomes solvable at M=2. I then plotted the absolute differences $|y_I^{MAP} - Y_I^{ML}|$ and $|y_F^{MAP} - Y_F^{ML}|$ and saw that they peak when M is small (reaching around 0.5 for M=2) and steadily decrease as more data are collected. By M≈20, the two estimates differ by less than 0.2. This shows that while the prior strongly regularizes the MAP solution under data scarcity, both MAP and ML estimates converge once enough measurements overwhelm the prior.

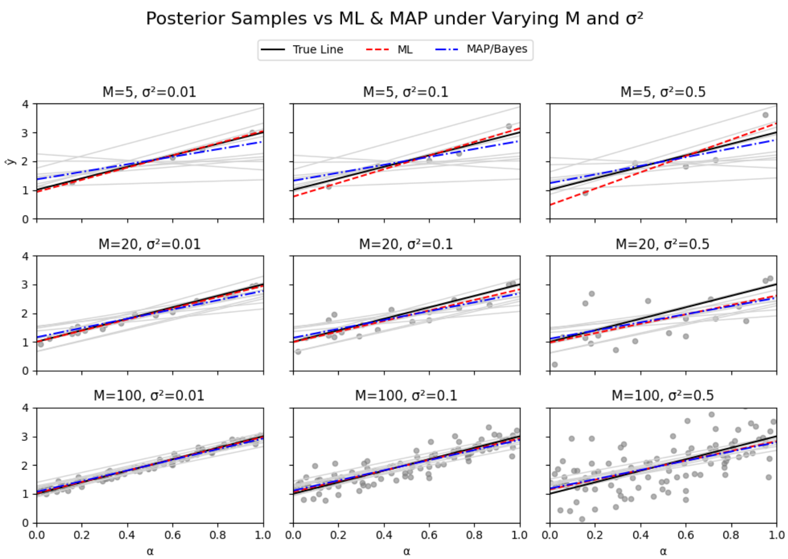


Figure 9

Week 9 -Question 4(f): full Bayesian estimation

Figure 9, shows the series of tests ran where I changed both the number of measurements (5, 20, 100) and how noisy those measurements were (low, medium, and high noise). When I only had a few points or the noise was high, the lines sampled from the Bayesian posterior spread out a lot, and the MAP fit was noticeably different from the ML fit—this shows that the full Bayesian view really helps by showing how uncertain we are about the line parameters. But once I had many measurements and the noise was low, all three lines (ML, MAP, and the Bayesian mean) lined up almost exactly, and the posterior samples clustered tightly. In those cases, knowing the full posterior doesn't add much beyond the single best estimate. In short, the Bayesian approach is most useful when data are scarce or noisy, and less critical when you have plenty of clean measurements.

Week 8 - Question 4(f): ML estimation

Figure 8 shows the series of tests I ran where I changed the number of measurements M (1, 2, 3, 5, 10, 20, 50, 100) and examined the resulting MLE line fit. When I had only one data point, the normal equations were singular, and no solution existed. From M=2 onward the solver succeeded, but with very few points (e.g.\ M=2–5) the fitted line was erratic and the absolute errors in $|y_I^{ML} - y_I^{True}|$ and $|y_F^{ML} - y_F^{True}|$ reached as high as 0.4. As I increased MM, the estimates stabilized: by M=10 both errors fell below 0.15, and beyond M=50 they hovered around 0.05–0.06. In short, at least two measurements are required to produce a valid MLE line, and adding more measurements steadily improves accuracy, with diminishing gains once M is large.

Effect of Measurement Noise -(Refers to: Q3(f), Q4(f) of Week 9 + Q4(g) of Week 8)

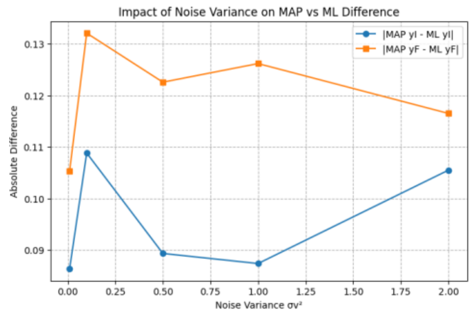


Figure 10

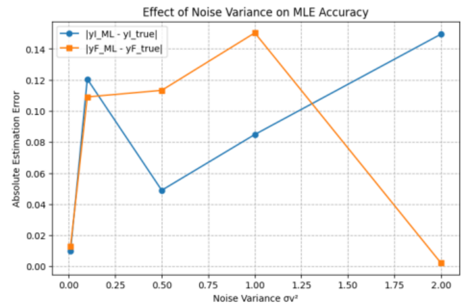


Figure 11

Week 9 - Question (3f): MAP Estimation

Figure 10 shows how the absolute difference between the MAP and ML estimates for both yI (blue circles) and yF (orange squares) varies as I increased the measurement noise variance from 0.01 up to 2.0, with M=50M held constant. I observed that when noise is very low (σ_v^2=0.01), the difference is small (around 0.08 for yI and 0.105 for yF). As I increased noise to moderate levels (σ_v^2= 0.1–1.0), the distance initially rose (peaking near 0.109 and 0.132 at σ_v^2= 0.1) and then dipped slightly. At the highest noise (σv2=2.0), the difference remains larger than in the low-noise case (about 0.105 and 0.117), showing that higher measurement noise amplifies the influence of the prior and widens the gap between MAP and ML. In summary, as measurement noise increases, MAP and ML diverge more, whereas under low noise they stay closer.

Week 9 -Question 4(f): full Bayesian estimation

The discussion for 4f of week 9 is under second section.

Week 8 - Question 4(g): ML estimation

Figure 11 shows how the absolute errors in y_I^{hat} (blue circles) and y_F^{hat} (orange squares) vary as I increase the noise variance σv2 from 0.01 to 2.0, with M=50 fixed. The ML point estimates themselves are obtained via the same closed-form solution, which does not explicitly depend on σv2; however, as the noise variance rises, the measurements become more scattered, and the resulting estimation error grows. Specifically, the error is minimal at very low noise (σv2=0.01), increases at moderate noise levels, and—though it dips at σv2=0.5 for y_F —reaches another peak near σv2=1.0. At very high noise (σv2=2.0), the error for y_F drops sharply , while the error for climbs to its highest. In summary, although the MLE formula remains the same, larger σv2 leads to less accurate line fits because it increases the variability in the data used by that formula.