CP2 Report

Collaboration Statement:

Professor Mike Hughes Office Hours

Total hours spent: 10

Links: [CP2 instructions] [Course collaboration policy]

Contents

1a

Given a dataset of size N, how do we score the model's predictions? Translate the provided starter code for the score function of the MAP estimator into a mathematical expression involving our probabilistic model. Provide a function in terms of parameters w_{MAP}, β and dataset $x_n, t_n{}_{n=1}^{N}$.

1a: Solution The log of likelihood formula as given in Bishop 3.11 is:

$$\begin{aligned}\ln p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) &= \sum_{n=1}^{N} \ln \mathcal{N}(t_n|\mathbf{w}^T\phi(\mathbf{x}_n), \beta^{-1}) \\ &= \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) - \beta\left(\frac{1}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^T\phi(\mathbf{x}_n))^2\right)\end{aligned}$$

The score formula divides the log likelihood by the N which represents the number of data points (observations) in the dataset:

$$= \frac{1}{N} \cdot \left(\frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) - \beta \cdot (\frac{1}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^T\phi(\mathbf{x}_n))^2)\right)$$

1b

Explain why different β values might be preferred by different model orders when using MAP. (Hint: Execute run_demo_MAP.py at 3 different β values: $\{0.1, 1, 10\}$. Look at the MAP estimator's scores on the training-set to see which β is preferred by which order. Then study the provided visualization to get intuition about why.).

1b Solution:

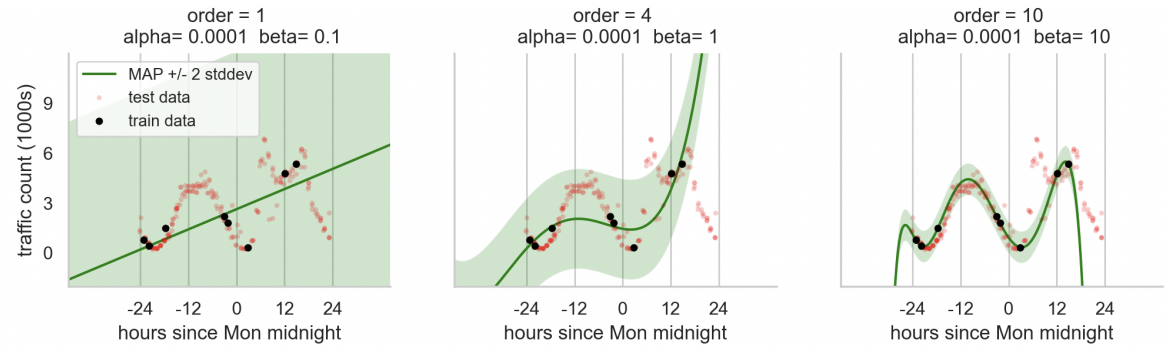

Figure 1: MAP estimator's scores on the training-set

Different β values might be prefered for different model orders when using MAP because the right choice of β helps managing the trade-off between fitting the data closely and not being influenced by random variations. (In other words, β controls how much emphasis the model puts on fitting exactly to the data points it's been trained on.)

Fitting Error Term =

$$-\beta \cdot (\frac{1}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^T\phi(\mathbf{x}_n))^2)$$

For,

1. **Low-Order Model (ex: order = 1 in the graph)** These models are less flexible and may not capture the underlying pattern in data well. A lower β would be preferred because it indicates higher noise variance and gives less weight to

the fitting error. This allows the model to fit the data more loosely and mitigate the high bias.

2. **High-Order Models (ex: order = 10 in the graph)** A higher beta value gives more weight to the fitting error, which is beneficial for more complex models that can capture more intricate patterns in data. For these models, a higher beta value penalizes any deviation from the actual target values and encourages the model to fit the data more closely (which is possible due to higher order), assuming there is less noise to account for.

1c:

Compare the visuals produced by run_demo_PPE.py to those from the MAP estimator. What do you notice about the width of the 2-stddev intervals of the predictions at x values far from the train data? Why does this suggest PPE might generalize better than MAP?

1c Solution:

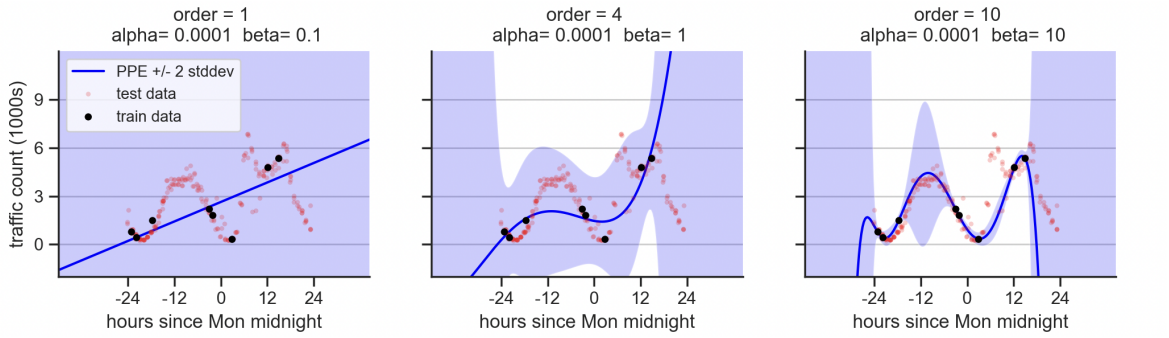

Figure 2: PPE estimator's scores on the training-set

In the PPE results, the 2-standard deviation intervals widen as we move from areas of dense training data to areas with few or no training data. This indicates that there is increasing uncertainty in regions where model has less information. On the other hand, in the MAP results, the confidence intervals tend to be more uniform regardless of whether we are close to or further from training data. This feature of increasing uncertainty of the graph at points that are away from the known data points suggest that PPE generalizes better than MAP. The reason is that PPE acknowledges that it is less sure about areas with less data points and is less likely to make overconfident predictions in those regions. This is potentially a more accurate representation of the data and can help prevent overfitting on test data.

2a:

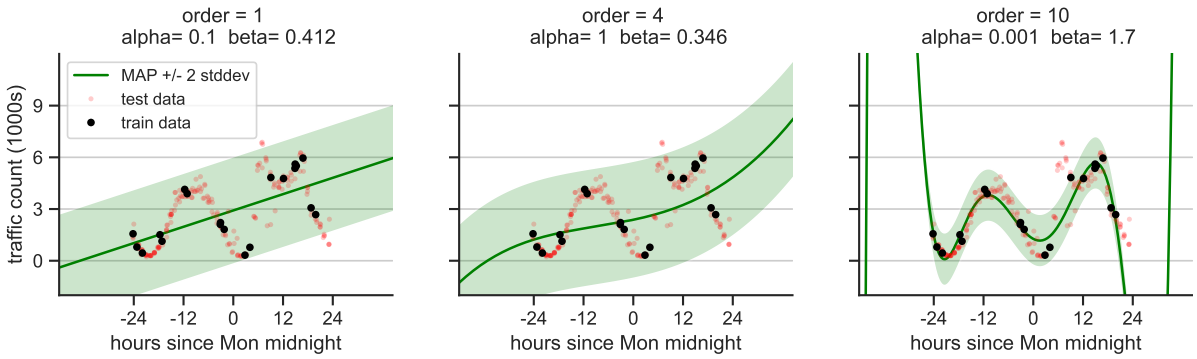

Fig. 2a: Probabilistic predictions of CV-selected models

2b

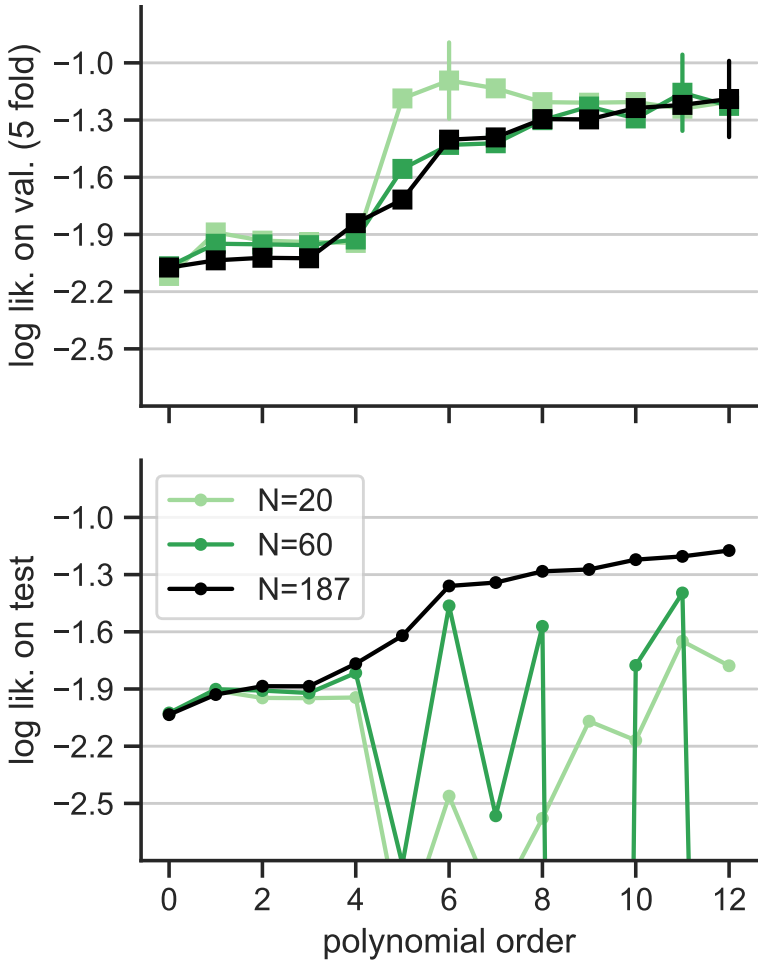

Fig. 2b: Model Score vs. Polynomial Order

3a

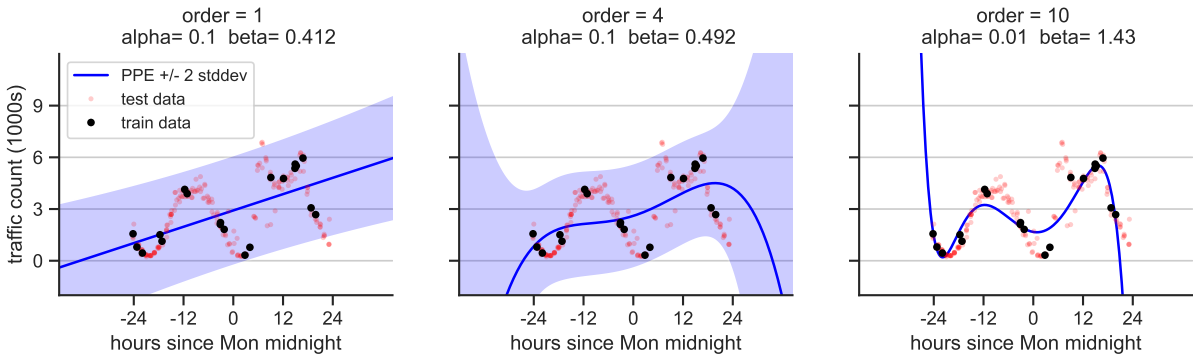

Fig. 3a: Probabilistic predictions of evidence-selected models

3b

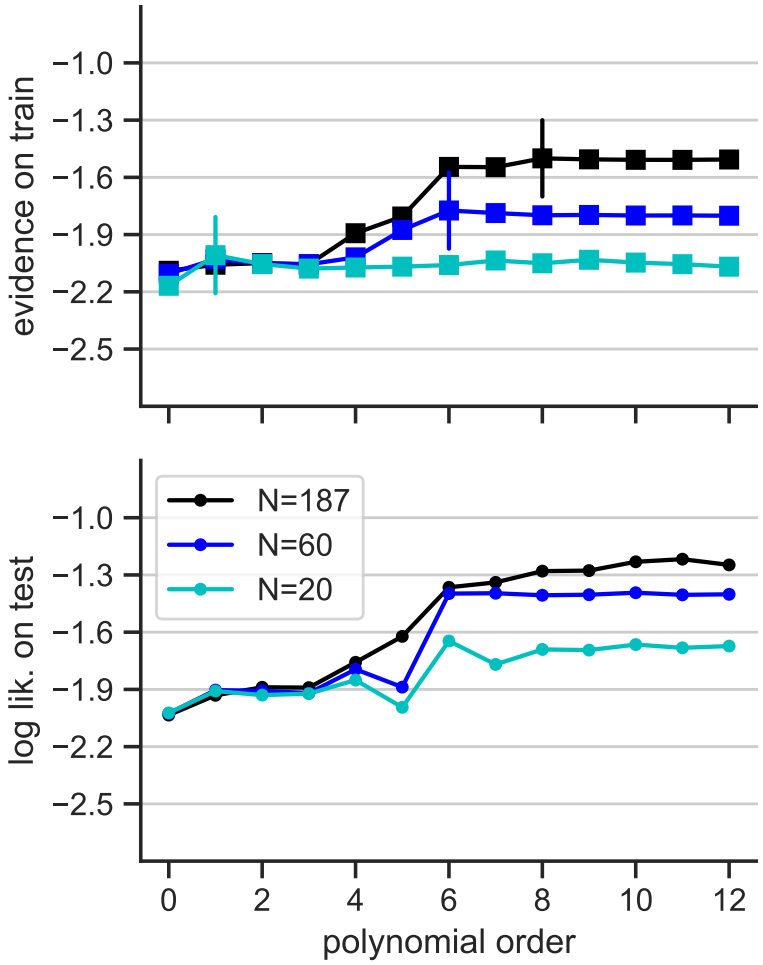

Fig. 3b: Model Evidence vs. Polynomial Order