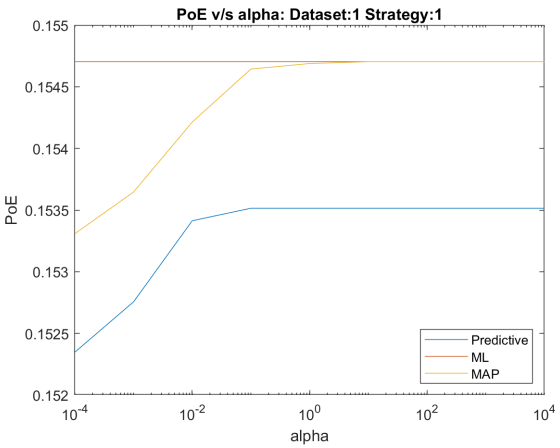
Bayesian Parameter Estimation

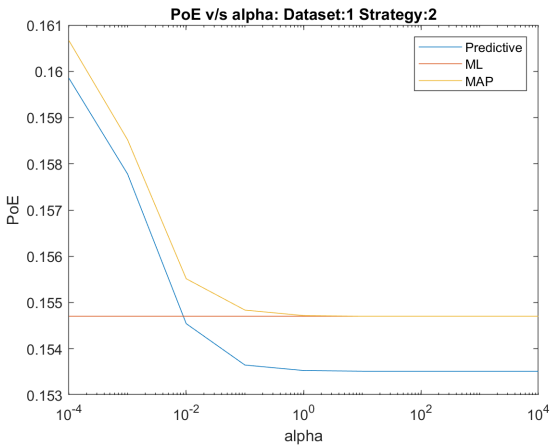
Saurabh Mirani

November 2019

1. The graphs are:

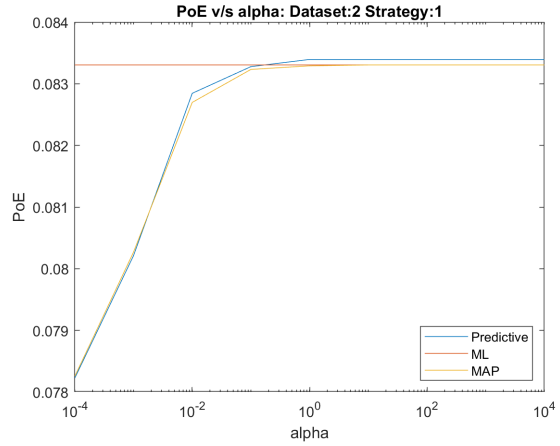


(a) Strategy 1

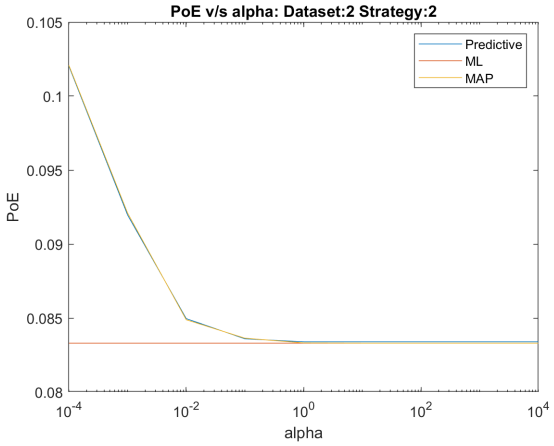


(b) Strategy 2

Figure 1: Datatset 1

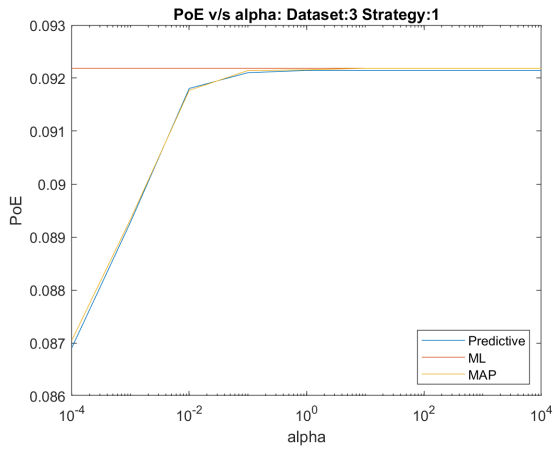


(a) Strategy 1

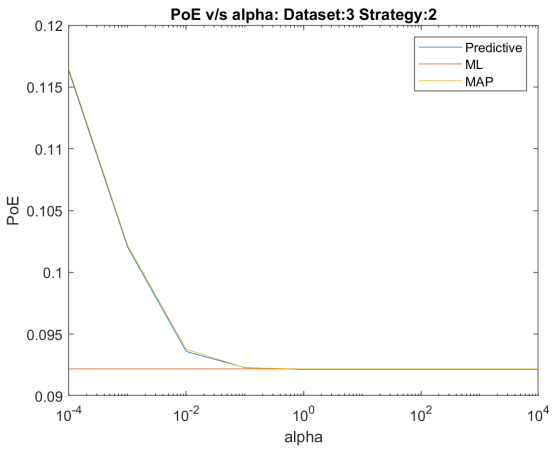


(b) Strategy 2

Figure 2: Datatset 2

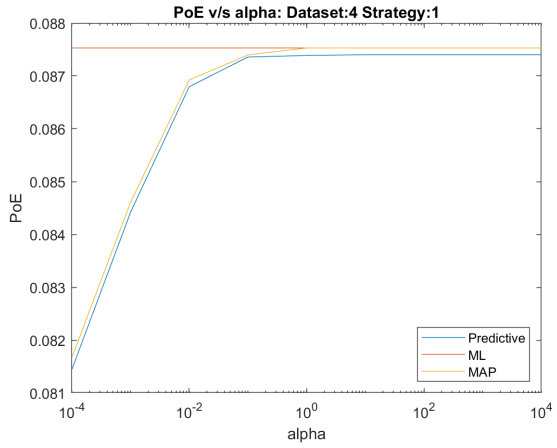


(a) Strategy 1

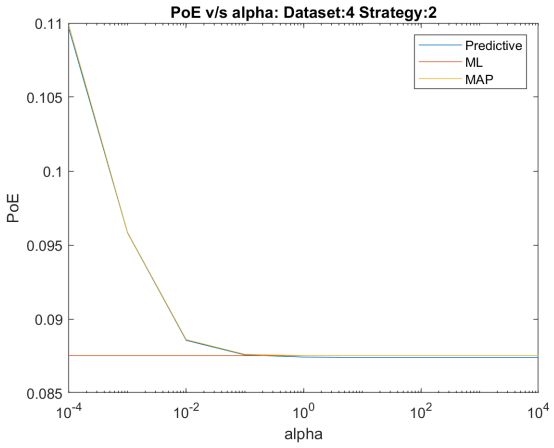


(b) Strategy 2

Figure 3: Datatset 3



(a) Strategy 1



(b) Strategy 2

Figure 4: Datatset 4

2. **The relative behavior of the three curves**

- The Predictive model almost always performs better than MAP. (some exceptions). Both Predictive and MAP perform better than or equal to ML for Strategy 1 and worse for Strategy 2.
- The ML curve was expected to be constant since it does no depend on the prior value. Hence, it does not change with the change in prior co-variance. The MAP and the Predictive distributions have same means but different co-variances.

$$ML : P_{x|T} = G(x, \hat{\mu_n}, \Sigma)$$

$$MAP : P_{x|T} = G(x, \mu_n, \Sigma)$$

$$Predictive : P_{x|T} = G(x, \mu_n, \Sigma + \Sigma_n)$$

$$where,\ \Sigma_n = \Sigma_o(\Sigma_o + \frac{1}{n}\Sigma_n)^{-1}\frac{1}{n}\Sigma_n$$

- Now, $\lim_{\alpha \to 0} \Sigma_n = 0$ and $\lim_{\alpha \to \infty} \Sigma_n = \frac{1}{n}\Sigma_n$
  Hence, the difference between the curves of MAP and Predictive will increase with increase in α.
- Now, $\lim_{\alpha \to 0} \mu_n = \mu_o$ and $\lim_{\alpha \to \infty} \mu_n = \hat{\mu_n}$
  Hence, as α increases MAP curve will converge to ML curve. The Predictive curve will also converge to ML curve but with a slight offset due to Σ_n term.
- This can also be understood intuitively. By increasing α we are trusting less on our prior and more on the data. This is what ML does. It does not include any prior. Hence, both MAP and Predictive tend to converge for higher α.
- Now, talking about the offset in prior, by making $\alpha \to \infty$ we are basically saying that our prior is non-informative,i.e. it has infinite variance. Hence, we get an offset from ML in the Predictive case.
- Now, talking about $\alpha \to 0$ case. If we have a good prior then the results will be better than those obtained by ML. This can be clearly seen in the graphs. Predictive almost always gives better results than MAP and MAP better than ML. (Strategy 1) {Explained more in Point - 4}

3. **How the behavior changes from dataset to dataset?**

   - As we go from dataset 1 to dataset 4, n increases. As n increases, less importance is given to prior and more importance is given to data. That is, data starts winning. Also as $\alpha \to \infty$ data is given more importance. It can be clearly seen that for higher values of α, the difference between Predictive & ML, MAP is huge for Dataset 1, considerable for Datatset 2, and almost negligible for Dataset 3 and 4.

4. **How all of the above change when Strategy 1 is replaced by Strategy 2?**

   - The only difference between strategy 1 and 2 is the prior mean. In strategy 1, different prior means are assigned to the two classes. However, in strategy 2, same prior means are assigned to both the classes. Strategy 1 is of-course better than strategy 2, since 2 classes will have different means. This can be clearly seen in the graphs. In strategy 2, Predictive and MAP perform worse than ML.
   - We have already seen that as $\alpha \to \infty$, prior is given less importance. Hence, Predictive and MAP both give approximately the same value as ML for higher $\alpha's$. Hence, we see an increasing trend in Strategy 1 while decreasing in Strategy 2.