Method 3

Tuesday, November 9, 2021 12:55 PM

GIVEN: $P_{X|\mu,\Sigma} = G(x, \mu, \Sigma)$ s.t. $\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \frac{1}{N}\sum_i x_i)(x_i - \frac{1}{N}\sum_i x_i)^T$

$\mu = ?$

$P_\mu(\mu) = G(\mu, \mu_0, \Sigma_0)$ s.t.

strategy 1: $\begin{cases} \text{cheetah: } \mu_0 = [1\ 0\ \cdots\ 0]^T \\ \text{grass: } \mu_0 = [3\ 0\ \cdots\ 0]^T \\ \Sigma_0 = \text{diag}(\alpha w_i) \text{ s.t. } w_i \text{ in Prior\_1.mat} \end{cases}$

strategy 2: $\begin{cases} \text{cheetah and grass: } \mu_0 = [2\ 0\ \cdots\ 0]^T \\ \Sigma_0 = \text{diag}(\alpha w_i) \text{ s.t. } w_i \text{ in Prior\_2.mat} \end{cases}$

A) training set D_1: for each class, compute · Σ = formula above where x are samples in D_1

· μ_1, · Σ_1 } s.t. $P_{\mu|T}(\mu|D_1) = G(\mu, \mu_1, \Sigma_1)$ $\Longrightarrow$ $\mu_1 = \Sigma_0(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}\underbrace{\mu_{ML}}_{\frac{1}{n}\sum_i x_i} + \frac{1}{n}\Sigma(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}\mu_0$

$\Sigma_1 = \Sigma_0(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}\frac{1}{n}\Sigma$

· parameters of $P_{X|T}(x|D_1)$

$\Longrightarrow P_{X|T}(x|D_1) = G(x, \mu_1, \Sigma + \Sigma_1)$

B) for ML procedure, we get: $P_{X|T}(x|D_1) = G(x, \mu_{ML}, \Sigma_{ML})$

$\mu_{ML} \downarrow \frac{1}{n}\sum_i x_i$ $\quad$ $\Sigma_{ML} \downarrow \Sigma$ from formula above

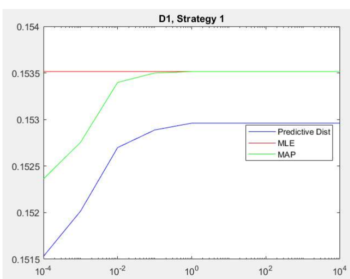
C) $P_{X|T}(x|D_1) = P_{X|\mu}(x|\mu_{MAP})$ $\Longrightarrow$ $\mu_{MAP} = \arg\max_\mu P_{\mu|T}(\mu|D_1)$

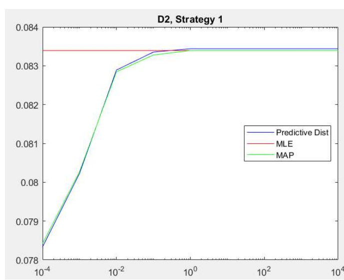
$= \arg\max_\mu G(\mu, \mu_1, \Sigma_1)$

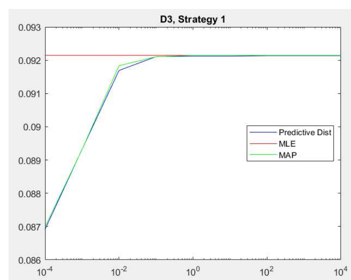
$= \mu_1$ by definition of Gaussian

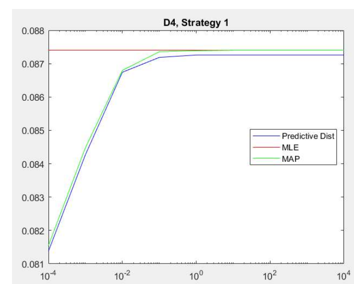
$= G(x, \mu = \mu_{MAP}, \Sigma)$

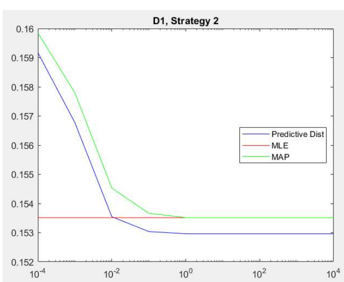
$= G(x, \mu_1, \Sigma)$

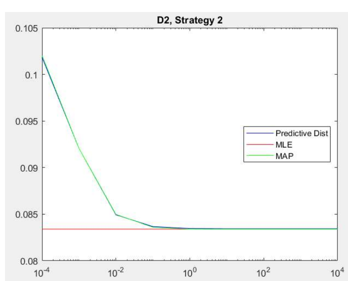


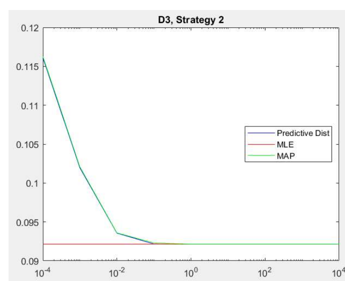


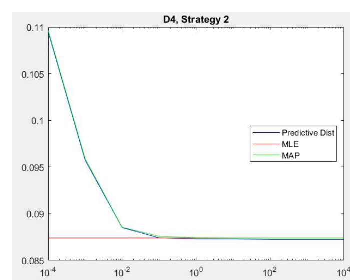












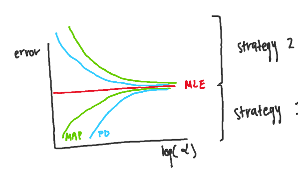
· MLE: μ_{ML} and Σ are not dependent on α, and hence the error is constant.

· MAP: μ_1 is dependent on α; when $\alpha \to \infty$, the diagonal elements of $\Sigma_0 \to \infty$ $\Rightarrow$ $\mu_1 = \underbrace{\Sigma_0(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}}_{\to 1}\mu_{ML} + \underbrace{\frac{1}{n}\Sigma(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}}_{\to 0}\mu_0 = \mu_{ML}$

· PD: $\mu_1 \to \mu_{ML}$ as $\alpha \to \infty$ as shown for MAP.

$\Sigma_1 \to \frac{1}{n}\Sigma$ as $\alpha \to \infty$, so the variance for the predictive distribution: $\Sigma + \Sigma_1 \to (1 + \frac{1}{n})\Sigma$.

Hence, for large n, this distribution is equal to the MAP (D_2, D_3, D_4). For smaller n (D_1), we have a higher variance.



· GENERAL: when $\alpha \to \infty$, $\Sigma_0 \to \infty$ meaning that the variance of the prior $\to \infty$ and hence the prior is not helpful in finding the solution, so we rely on the data. We rely on prior when $\alpha \to 0$ and n is small

· in strategy 1, the prior is accurate meaning the more we rely on it ($\alpha \to 0$), the better the error.

· in strategy 2, the prior is non-informative meaning the less we rely on it ($\alpha \to \infty$), the better the error.

· D_1 generates the highest error (smallest n), and D_2, D_3, D_4 are all about the same magnitude