STATISTICAL MACHINE LEARNING (WS2024)
EM ALGORITHM AND BAYESIAN LEARNING

Assignment 1. Let us consider a mixture of distributions from an exponential family, i.e.

$$p(x) = \sum_{k=1}^{K} \pi_k e^{\langle \phi(x), \eta_k \rangle - A(\eta_k)}$$

where $\eta = (\eta_1, \dots, \eta_K)$ is the tuple of natural parameters and $\pi = (\pi_1, \dots, \pi_K)$ is the tuple of mixture weights. Suppose you want to estimate η and π from a training set $\mathcal{T}^m = \{x^j \mid j = 1, \dots, m\}$ by using the EM algorithm. In the E-step you will need to compute the optimal auxiliary variables

$$\alpha_x(k) = p\big(k \,|\, x; \eta^{(t)}, \pi^{(t)}\big), \quad \forall x \in \mathcal{T}^m$$

for the current estimate of η and π. In the M-step you will need to solve the optimisation task

$$\frac{1}{m} \sum_{x \in \mathcal{T}^m} \sum_{k=1}^{K} \alpha_x(k) \Big[\langle \phi(x), \eta_k \rangle - A(\eta_k) + \log \pi_k \Big] \to \max_{\eta, \pi}$$

a) Show that the task decomposes into independent optimisation tasks for η and π. Find the optimal tuple of mixture weights π.

b) Show that the optimisation task w.r.t. η further decomposes into independent tasks for each η_k. Show that each of them is an ML estimate for the respective η_k with the statistics

$$\psi_k = \frac{\sum_{x \in \mathcal{T}^m} \alpha_x(k) \phi(x)}{\sum_{x' \in \mathcal{T}^m} \alpha_{x'}(k)}.$$

Assignment 2. Consider a training set $\mathcal{T}^m = ((x^i, y^i) \in \mathcal{X} \times \mathcal{Y} \mid i = 1, \dots, m)$, where the samples are i.i.d. and drawn from a distribution $p_{\text{tr}}(x, y) = p(x \mid y)\, p_{\text{tr}}(y)$. Using this training set, a posterior estimate $\hat{p}_{\text{tr}}(y \mid x)$ is learned, and a plugin Bayes classifier is constructed as follows:

$$\hat{h}(x) = \arg\min_{y \in \mathcal{Y}} \sum_{y' \in \mathcal{Y}} \hat{p}_{\text{tr}}(y' \mid x) \ell(y, y') \,, \tag{1}$$

where $\ell \colon \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$ is a given loss function.

At deployment, however, the data distribution changes such that the prior shifts from the training prior $p_{\text{tr}}(y)$ to a new deployment prior $p_{\text{de}}(y)$. The deployment data is thus generated from $p_{\text{de}}(x, y) = p(x \mid y)\, p_{\text{de}}(y)$, where $p_{\text{de}}(y) \neq p_{\text{tr}}(y)$.

Your are provided with estimates of both training prior $\hat{p}_{\mathrm{tr}}(y)$ and the deployment prior $\hat{p}_{\mathrm{de}}(y)$. Your task is to show how to adjust the plugin Bayes classier $\hat{h}(x)$ to efficiently handle the deployment data, considering the prior shift.

Assignment 3. Building on the setup described in Assignment 2, consider a scenario where you no longer have an estimate of the deployment prior $p_{\mathrm{de}}(y)$. Instead, your are provided with an unlabeled data $\mathcal{S}^n = (x^i \in \mathcal{X} \mid i = 1, \ldots, n)$, where the samples are i.i.d. and drawn from the deployment distribution

$$p_{\mathrm{de}}(x) = \sum_{y \in \mathcal{Y}} p_{\mathrm{de}}(x, y) \,.$$

Your task is to derive a method for estimating the deployment prior $p_{\mathrm{de}}(y)$ using the unlabeled samples $\mathcal{S}^n$. Specifically, you are required to use the EM algorithm to derive the estimation procedure.

Assignment 4. Given a small training set $\mathcal{T}^m = (x_i \in \mathbb{R} \mid i = 1, \ldots, m)$ we want to estimate the mean of a normal distribution $\mathcal{N}(\mu, 1)$. We know that the unknown μ is close to μ_0. Therefore, we want to apply Bayesian inference and set the prior distribution for μ to be a normal distribution centred at μ_0, i.e. $p(\mu) = \mathcal{N}(\mu_0, 1)$.

Show that the posterior distribution $p(\mu \,|\, \mathcal{T}^m) \propto p(\mathcal{T}^m \,|\, \mu)\; p(\mu)$ is also a Gaussian. Find its center (expectation).