Homework 3

Deadline: Thursday, Nov. 28, at 11:59pm.

Submission: You need to submit your answers to all 3 questions through MarkUs[1] as a PDF file titled `hw5_writeup.pdf`. You can produce the file however you like (e.g. LaTeX, Microsoft Word, scanner), as long as it is readable.

Late Submission: 10% of the marks will be deducted for each day late, up to a maximum of 3 days. After that, no submissions will be accepted.

Collaboration: Homeworks are individual work. See the course web page for detailed policies.

1. [**5 points**] **EM for Probabilistic PCA.** In lecture, we covered the EM algorithm applied to mixture of Gaussians models. In this question, we'll look at another interesting example of EM but where the latent variables are continuous: probabilistic PCA. This is a model very similar in spirit to PCA: we have data in a high-dimensional space, and we'd like to summarize it with a lower-dimensional representation. Unlike ordinary PCA, we formulate the problem in terms of a probabilistic model. We assume the latent code vector $\mathbf{z}$ is drawn from a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, and that the observations are drawn from a spherical Gaussian whose mean is a linear function of $\mathbf{z}$. We'll consider the slightly simplified case of scalar-valued z (i.e. only one principal component). The probabilistic model is given by:

$$z \sim \mathcal{N}(0, 1)$$
$$\mathbf{x} \,|\, z \sim \mathcal{N}(z\mathbf{u}, \sigma^2 \mathbf{I}),$$

where σ^2 is the noise variance (which we assume to be fixed) and $\mathbf{u}$ is a parameter vector (which, intuitively, should correspond to the top principal component). Note that the observation model can be written in terms of coordinates:

$$x_j \,|\, z \sim \mathcal{N}(zu_j, \sigma^2).$$

We have a set of observations $\{\mathbf{x}^{(i)}\}_{i=1}^N$, and z is a latent variable, analogous to the mixture component in a mixture-of-Gaussians model.

In this question, you'll derive both the E-step and the M-step for the EM algorithm.

   (a) **E-step (2 points).** In this step, your job is to calculate the statistics of the posterior distribution $q(z) = p(z \,|\, \mathbf{x})$ which you'll need for the M-step. In particular, your job is to find formulas for the (univariate) statistics:

$$m = \mathbb{E}[z \,|\, \mathbf{x}] =$$
$$s = \mathbb{E}[z^2 \,|\, \mathbf{x}] =$$

   Tips:

   - First determine the conditional distribution $p(z \,|\, \mathbf{x})$ using the Gaussian conditioning formulas from the Appendix.. To help you check your work: $p(z \,|\, \mathbf{x})$ is a univariate Gaussian distribution whose mean is a linear function of $\mathbf{x}$, and whose variance does not depend on $\mathbf{x}$.

[1] `https://markus.teach.cs.toronto.edu/csc2515-2019-09`

- Once you've determined the conditional distribution (and hence the posterior mean and variance), use the fact that $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$ for any random variable Y.

(b) **M-step (3 points).** In this step, we need to re-estimate the parameters, which consist of the vector $\mathbf{u}$. (Recall that we're treating σ as fixed.) Your job is to derive a formula for $\mathbf{u}_{\text{new}}$ that maximizes the expected log-likelihood, i.e.,

$$\mathbf{u}_{\text{new}} \leftarrow \arg\max_{\mathbf{u}} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{q(z^{(i)})}[\log p(z^{(i)}, \mathbf{x}^{(i)})].$$

(Recall that $q(z)$ is the distribution computed in part (a).) This is the new estimate obtained by the EM procedure, and will be used again in the next iteration of the E-step. Your answer should be given in terms of the $m^{(i)}$ and $s^{(i)}$ from the previous part. (I.e., you don't need to expand out the formulas for $m^{(i)}$ and $s^{(i)}$ in this step, because if you were implementing this algorithm, you'd use the values $m^{(i)}$ and $s^{(i)}$ that you previously computed.)

Tips:

- First expand out $\log p(z^{(i)}, \mathbf{x}^{(i)})$. You'll find that a lot of the terms don't depend on $\mathbf{u}$ and can therefore be dropped.
- Apply linearity of expectation. You should wind up with terms proportional to $\mathbb{E}_{q(z^{(i)})}[z^{(i)}]$ and $\mathbb{E}_{q(z^{(i)}}[[z^{(i)}]^2]$. Replace these expectations with $m^{(i)}$ and $s^{(i)}$. You should get an equation that does not mention $z^{(i)}$. (If you don't wind up with terms of this form, then see if there's some way you can simplify $\log p(z^{(i)}, \mathbf{x}^{(i)})$.
- In order to find the maximum likelihood parameter $\mathbf{u}_{\text{new}}$, you need to determine the gradient with respect to $\mathbf{u}$, set it to zero, and solve for $\mathbf{u}_{\text{new}}$.

2. **[2 points] Contraction Maps.** In lecture, we showed that the optimal Bellman backup operator is a contraction map, and hence that value iteration converges to the optimal Q-function Q^*. Now consider the problem of *policy evaluation*, i.e. finding the Q-function Q^π for a given (stochastic) policy π. Since Q^π is characterized by the fixed-point equation $T^\pi Q^\pi = Q^\pi$, we can repeatedly apply the update

$$Q_{k+1} \leftarrow T^\pi Q_k,$$

which can be written out in full as:

$$Q_{k+1}(s, a) \leftarrow r(s, a) + \gamma \sum_{s'} \mathcal{P}(s' \mid a, s) \sum_{a'} \pi(a' \mid s') \, Q_k(s', a').$$

Show that the Bellman backup operator T^π is a contraction map in the $\|\cdot\|_\infty$ norm. Your proof will probably look very similar to the one from Slide 30 of Lecture 10, but be sure to justify each step.

3. **[3 points] Q-Learning.** In lecture, we made the claim that Q-learning only converges to the optimal Q-function if the agent follows an exploration-encouraging strategy such as ε-greedy. Your job is to give a counterexample to show that exploration is necessary. I.e., you will show that Q-learning might get stuck with a suboptimal Q-function if it always chooses $\pi(s) = \arg\max_a Q(s, a)$.

Consider an MDP with two states s_1 and s_2, and two actions, Stay and Switch. The environment is deterministic. If the agent chooses Stay, then it stays in the current state (i.e. $S_{t+1} = S_t$), while if it chooses Switch, it switches to the other state (i.e., if it's in s_1, it transitions to s_2, and vice versa). The reward function is given by:

$$r(S, A) = \begin{cases} 1 & \text{if } S = s_1 \\ 2 & \text{if } S = s_2 \end{cases}$$

The discount factor is $\gamma = 0.9$.

(a) **(1 point)** Determine the optimal policy and the Q-function for the optimal policy. You should give the Q-function as a table. You don't need to show your work or justify your answer for this part.

(b) **(2 points)** Now suppose we apply Q-learning, except that instead of the ε-greedy policy, the agent follows the greedy policy which always chooses $\pi(s) = \arg\max_a Q(s, a)$. Assume the agent starts in state $S_0 = s_1$. Give an example of a Q-function that is in equilibrium (i.e. it will never change after the Q-learning update rule is applied), but which results in a suboptimal policy. (You should specify the Q-function as a table.) Justify your answer.

Appendix: Some Properties of Gaussians

Consider a multivariate Gaussian random variable $\mathbf{z}$ with the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. I.e.,

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Now consider another Gaussian random variable $\mathbf{x}$, whose mean is an affine function of $\mathbf{z}$ (in the form to be clear soon), and its covariance $\mathbf{S}$ is independent of $\mathbf{z}$. The conditional distribution of $\mathbf{x}$ given $\mathbf{z}$ is

$$p(\mathbf{x} \,|\, \mathbf{z}) = \mathcal{N}(\mathbf{x} \,|\, \mathbf{A}\mathbf{z} + \mathbf{b}, \mathbf{S}).$$

Here the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are of appropriate dimensions.

In some problems, we are interested in knowing the distribution of $\mathbf{z}$ given $\mathbf{x}$, or the marginal distribution of $\mathbf{x}$. One can apply Bayes' rule to find the conditional distribution $p(\mathbf{z} \,|\, \mathbf{x})$. After some calculations, we can obtain the following useful formulae:

$$\begin{aligned} p(\mathbf{x}) &= \mathcal{N}\left(\mathbf{x} \,|\, \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top + \mathbf{S}\right) \\ p(\mathbf{z} \,|\, \mathbf{x}) &= \mathcal{N}\left(\mathbf{z} \,|\, \mathbf{C}(\mathbf{A}^\top\mathbf{S}^{-1}(\mathbf{x} - \mathbf{b}) + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}), \mathbf{C}\right) \end{aligned}$$

with

$$\mathbf{C} = (\boldsymbol{\Sigma}^{-1} + \mathbf{A}^\top\mathbf{S}^{-1}\mathbf{A})^{-1}.$$

You may also find it helpful to read Section 2.3 of Bishop.