

HW 2

Due: Friday 10/24/2025 @ 11:59pm EST

Disclaimer

I encourage you to work together, I am a firm believer that we are at our best (and learn better) when we communicate with our peers. Perspective is incredibly important when it comes to solving problems, and sometimes it takes talking to other humans (or rubber ducks in the case of programmers) to gain a perspective we normally would not be able to achieve on our own. The only thing I ask is that you report who you work with: this is **not** to punish anyone, but instead will help me figure out what topics I need to spend extra time on/who to help. When you turn in your solution (please use some form of typesetting: do **NOT** turn in handwritten solutions), please note who you worked with.

Question 1: Matrix Derivatives (30 points)

Differentiating scalars is great, but what happens if our objective function is written using linear algebra? The objective may still be differentiable, and so to optimize it we would still need to differentiate it. But how can you differentiate linear algebra equations? In this question you will be deriving some of the derivative rules for matrix equations.

Remember that matrix equations are used to group together lots of smaller, scalar equations. Differentiation also follows a similar rule: until you know the matrix derivative rules as well as the scalar ones, the procedure for differentiating a matrix equation is as follows:

1. expand the matrix equation into its scalar form. This equation may be huge and will probably involve lots of summations.
2. differentiate the giant scalar equation.
3. repackage the giant derivative equation back into matrix form.

Please follow this procedure to derive the matrix derivative rules for the following matrix equations:

1. (6 points) $\frac{\partial \mathbf{AB}}{\partial \mathbf{A}}$
2. (6 points) $\frac{\partial \mathbf{AB}}{\partial \mathbf{B}}$
3. (6 points) $\frac{\partial \|\vec{x}\|_p^p}{\partial \vec{x}}$
4. (6 points) $\frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial \vec{x}}$
5. (6 points) $\frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial \mathbf{A}}$

Question 2: Multivariate Gaussian Mixture Models (35 points)

A Gaussian mixture model is an instance of a very specific application of expectation maximization. The problem in lecture we've been applying expectation maximization to has been placing examples into exactly one of k distinct buckets, where each bucket is controlled with a (potentially different shape from the others) probability distribution. A Gaussian mixture model makes a further assumption to this problem: it assumes that every bucket is controlled by a separate gaussian distribution. The pdf for a 1-dimensional gaussian is:

$$Pr[x; \mu, \sigma^2] = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

In your labs you have solved this problem for a 1-dimensional gaussian distribution by first deriving the log-likelihood of a set of samples (denoted as X) as:

$$\mathcal{LL}(X, Z; \theta, \pi) = \sum_{x_i \in X} \sum_{j=1}^k z_{ij} \left(\log(\pi_j) + \log(Pr[x_i; \mu_j, \sigma_j^2]) \right)$$

and then deriving:

$$Q(X; \theta, \pi) := \mathbb{E}_Z[\mathcal{LL}(X, Z; \theta, \pi)] = \sum_{x_i \in X} \sum_{j=1}^k Pr[z_{ij} = 1 | x_i] \left(\log(\pi_j) + \log(Pr[x_i; \mu_j, \sigma_j^2]) \right)$$

In lecture we already applied the e-step of expectation maximization to maximize Q for the alignment (i.e. maximizing $Pr[z_{ij} = 1 | x_i]$). The solution was to calculate:

$$\forall x_i \in X, \forall j \in [k] \quad \gamma_{ij} = \frac{Pr[x_i; \mu_j, \sigma_j^2] \pi_j}{\sum_{m=1}^k Pr[x_i; \mu_m, \sigma_m^2] \pi_m}$$

After plugging this solution back into Q , we arrived at the m-step of expectation maximization where we needed to figure out π and μ, σ^2 for every cluster. We took our modified Q :

$$Q(X; \theta, \pi) = \sum_{x_i \in X} \sum_{j=1}^k \gamma_{ij} \left(\log(\pi_j) + \log(Pr[x_i; \mu_j, \sigma_j^2]) \right) \quad \text{s.t.} \quad \sum_{j=1}^k \pi_j = 1$$

found the equations for $\frac{\partial Q}{\partial \pi_m}, \frac{\partial Q}{\partial \mu_m}, \frac{\partial Q}{\partial \sigma_m^2}$, set them equal to zero and solved. We arrived at the following solutions:

$$\begin{aligned} \pi_m^* &= \frac{1}{n} \sum_{x_i \in X} \gamma_{im} \\ \mu_m^* &= \frac{\sum_{x_i \in X} \gamma_{im} x_i}{\sum_{x_i \in X} \gamma_{im}} \\ (\sigma^2)_m^* &= \frac{\sum_{x_i \in X} \gamma_{im} (x_i - \mu_m^*)^2}{\sum_{x_i \in X} \gamma_{im}} \end{aligned}$$

In this question, what if each sample is **multi-dimensional** instead of a scalar? In this case each sample becomes a **vector** in d -dimensional space and will be denoted as \vec{x} . Gaussian exist for multi-dimensional samples, but the pdf changes to now have a mean vector $\vec{\mu}$ and a covariance matrix Σ . The new pdf looks like:

$$Pr[\vec{x}; \vec{\mu}, \Sigma] = \frac{e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})}}{\sqrt{(2\pi)^k |\Sigma|}}$$

where Σ^{-1} is the inverse, $|\Sigma|$ is the determinant of covariance matrix Σ , and k is the dimensionality of \vec{x} . Most of the math stays the same (γ_{ij} and π_m^* stay the same) and we arrive at the following Q :

$$Q(X; \theta, \pi) = \sum_{\vec{x}_i \in X} \sum_{j=1}^k \gamma_{ij} \left(\log(\pi_j) + \log(Pr[\vec{x}_i; \vec{\mu}_j, \Sigma_j]) \right) \quad \text{s.t.} \quad \sum_{j=1}^k \pi_j = 1$$

Please find the MLE estimates for:

1. (15 points) $\vec{\mu}_m^*$
2. (20 points) Σ_m^*

Question 3: Singular Value Decomposition) (35 points)

Consider matrix \mathbf{A} and its svd:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

1. (15 points) Show that for a square, symmetric matrix \mathbf{M} , any two eigenvectors \vec{v}_1, \vec{v}_2 with distinct eigenvalues λ_1, λ_2 are orthogonal (i.e. the inner product $\langle \vec{v}_1, \vec{v}_2 \rangle$ is 0). You will need the following lemma:

$$\text{for any matrix } \mathbf{M} \text{ and vectors } \vec{v}, \vec{w}, \langle \mathbf{M}\vec{v}, \vec{w} \rangle = \langle \vec{v}, \mathbf{M}^T \vec{w} \rangle$$

2. (10 points) Show that $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ are symmetric.
3. (5 points) Derive an expression that relates singular vectors and singular values of \mathbf{A} to the eigenvectors and eigenvalues of $\mathbf{A}\mathbf{A}^T$
4. (5 points) Derive an expression that relates singular vectors and singular values of \mathbf{A} to the eigenvectors and eigenvalues of $\mathbf{A}^T\mathbf{A}$

Extra Credit: Las-Vegas Johnson-Lindenstrauss (25 points)

In lecture we talked about the Johnson-Lindenstrauss lemma and how it gives us a Monte-Carlo algorithm for dimensionality reduction. In lab we converted the Monte-Carlo Johnson-Lindenstrauss algorithm into a Las-Vegas algorithm by running the Monte-Carlo version until it succeeded. Derive an expression on the expected number of iterations that the Las-Vegas Johnson-Lindenstrauss algorithm will need to execute before it terminates.