HWK 8

1 Recall that the trace) of a square matrix $A$, denoted $\text{tr}(A)$, is the sum of its diagonal entries.
   a) Show that, for any $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times n}$, it holds that $\text{tr}(AB) = \text{tr}(BA)$.
   b) Use a) to show that more generally $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$ for any matrices $A, B, C$ for which $AB, BC$ and $CA$ are well-defined.
   c) Show that, for any $A \in \mathbb{R}^{n \times m}$, $\|A\|^2_F = \text{tr}(A^T A)$.
   d) For a matrix $A = (a_{i,j})_{i,j} \in \mathbb{R}^{n \times m}$, the vectorization of $A$ is the following vector
      \[ \text{vec}(A) = (a_{1,1}, \ldots, a_{n,1}, a_{1,2}, \ldots, a_{n,2}, \ldots, a_{1,m}, \ldots, a_{n,m}) \]
      that is, it is obtained by stacking the columns of the matrix on top of one another. Show that, for any $A, B \in \mathbb{R}^{n \times m}$, it holds that $\text{tr}(A^T B) = \text{vec}(A)^T \text{vec}(B)$.

2 Let $A = \sum_{j=1}^r \sigma_j u_j v_j^T$ be an SVD of $A \in \mathbb{R}^{n \times m}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.
   Define
   \[ B = A - \sigma_1 u_1 v_1^T. \]
   Show that
   \[ v_2 \in \arg\max\{\|Bv\| : \|v\| = 1\}. \]

3 Let $A \in \mathbb{R}^{n \times n}$ be a square matrix with full SVD $A = U \Sigma V^T$.
   a) Justify the following formula
      \[ A = (UV^T)(V\Sigma V^T). \]
   b) Let
      \[ Q = UV^T, \quad S = V\Sigma V^T. \]
   Show that $Q$ is orthogonal and that $S$ is positive semidefinite. A factorization of the form $A = QS$ is called a polar decomposition.

4 Assume that, for each $i$, $p_{\theta_i}$ is a univariate Gaussian with mean $\theta_i = x_i^T \mathbf{w}$ and known variance $\sigma_i^2$. Show that the maximum likelihood estimator of $\mathbf{w}$ solves the weighted least squares problem, as defined in a previous assignment.
5 a) Show that the exponential family form of the Poisson distribution with mean \( \lambda \) has sufficient statistic \( \phi(y) = y \) and natural parameter \( \theta = \log \lambda \).

b) In Poisson regression, we assume that \( p_\theta(y) \) is Poisson with \( \theta = x^T w \). Compute the gradient and Hessian of the minus log-likelihood in this case.