HWK 8

1 Recall that the trace) of a square matrix A, denoted $\mathrm{tr}(A)$, 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 $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

b) Use a) to show that more generally $\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$ for any matrices A, B, C for which AB, BC and CA are well-defined.

c) Show that, for any $A \in \mathbb{R}^{n \times m}$, $\|A\|_F^2 = \mathrm{tr}(A^T A)$.

d) For a matrix $A = (a_{i,j})_{i,j} \in \mathbb{R}^{n \times m}$, the vectorization of A is the following vector

$$\mathrm{vec}(A) = (a_{1,1}, \ldots, a_{n,1}, a_{1,2}, \ldots, a_{n,2}, \ldots, a_{1,m}, \ldots, a_{n,m})$$

that is, it is obtained by stacking the columns of the matrix on top of one another. Show that, for any $A, B \in \mathbb{R}^{n \times n}$, it holds that $\mathrm{tr}(A^T B) = \mathrm{vec}(A)^T \mathrm{vec}(B)$.

2 Let $A = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T$ be an SVD of $A \in \mathbb{R}^{n \times m}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Define

$$B = A - \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T.$$

Show that

$$\mathbf{v}_2 \in \arg\max\{\|B\mathbf{v}\| : \|\mathbf{v}\| = 1\}.$$

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

$$Q = UV^T, \qquad S = V \Sigma V^T.$$

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_{θ_i} is a univariate Gaussian with mean $\theta_i = \mathbf{x}_i^T \mathbf{w}$ and known variance σ_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 λ 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 $\theta = \mathbf{x}^T \mathbf{w}$. Compute the gradient and Hessian of the minus log-likelihood in this case.