Spring 2018

1 Trace Derivatives

- (a) Let P be a $p \times q$ matrix and Q be a $q \times p$ matrix. Compute $\frac{\partial \text{trace}(PQ)}{\partial P}$.
- (b) Let \mathbf{P} be a $p \times q$ matrix and \mathbf{Q} be a $q \times q$ matrix. Compute $\frac{\partial \operatorname{trace}(\mathbf{PQP}^{\top})}{\partial \mathbf{P}}$ at $\mathbf{P} = \mathbf{U}$.

2 Unitary invariance

- (a) Prove that the regular Euclidean norm (also called the ℓ^2 -norm) is unitary invariant; in other words, the ℓ^2 -norm of a vector is the same, regardless of how you apply a rigid linear transformation to the vector (i.e., rotate or reflect). Note that rigid linear transformation of a vector $\mathbf{v} \in \mathbb{R}^d$ means multiplying by an orthogonal $\mathbf{U} \in \mathbb{R}^{d \times d}$.
- (b) Now show that the Frobenius norm of matrix \mathbf{A} is unitary invariant. The Frobenius norm is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j=1}^n |a_{ij}|^2} = \sqrt{tr(\mathbf{A}^\top \mathbf{A})}$.
- 3 Least Squares (using vector calculus)
- (a) In ordinary least-squares linear regression, we typically have n > d so that there is no w such that $\mathbf{X}\mathbf{w} = \mathbf{y}$ (these are typically overdetermined systems too many equations given the number of unknowns). Hence, we need to find an approximate solution to this problem. The residual vector will be $\mathbf{r} = \mathbf{X}\mathbf{w} \mathbf{y}$ and we want to make it as small as possible. The most common case is to measure the residual error with the standard Euclidean ℓ^2 -norm. So the problem becomes:

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$$

Where $\mathbf{X} \in \mathbb{R}^{n \times d}$, $\mathbf{w} \in \mathbb{R}^d$, $\mathbf{y} \in \mathbb{R}^n$. Derive using vector calculus an expression for an optimal estimate for \mathbf{w} for this problem assuming \mathbf{X} is full rank.

- (b) How do we know that $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ is invertible?
- (c) What should we do if X is not full rank?