STA9890 - Test 1 - Formula Sheet

Linear Algebra:

- A *n-vector* is an ordered set of n (real) numbers: $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, with addition $\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$ and vector (inner / dot) product: $\boldsymbol{x} \bullet \boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^n x_i y_i$
- Vector norms: $\|\boldsymbol{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$ with $\|\boldsymbol{x}\|_\infty = \max_i\{|x_i|\}$ and $\|\boldsymbol{x}\|_0 =$ Number of non-zero elements of $\boldsymbol{x}$
- An $m \times n$ matrix is a 2D array of real numbers with m rows and n columns:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

- A matrix-vector product takes an n-vector as input and gives an m-vector as output:

$$\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}_{1\cdot} \bullet \boldsymbol{x}, \boldsymbol{A}_{2\cdot}\boldsymbol{x}, \dots, \boldsymbol{A}_{m\cdot} \bullet \boldsymbol{x}) \in \mathbb{R}^m$$

- We can multiply an $m \times n$ matrix with an $n \times p$ matrix - note that the 'inner' dimensions must match:

$$\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{A}_{1\cdot} \bullet \boldsymbol{B}_{\cdot 1} & \boldsymbol{A}_{1\cdot} \bullet \boldsymbol{B}_{\cdot 2} & \dots & \boldsymbol{A}_{1\cdot} \bullet \boldsymbol{B}_{\cdot n} \\ \boldsymbol{A}_{2\cdot} \bullet \boldsymbol{B}_{\cdot 1} & \boldsymbol{A}_{2\cdot} \bullet \boldsymbol{B}_{\cdot 2} & \dots & \boldsymbol{A}_{2\cdot} \bullet \boldsymbol{B}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{m\cdot} \bullet \boldsymbol{B}_{\cdot 1} & \boldsymbol{A}_{m\cdot} \bullet \boldsymbol{B}_{\cdot 2} & \dots & \boldsymbol{A}_{m\cdot} \bullet \boldsymbol{B}_{\cdot n} \end{pmatrix} \in \mathbb{R}^{m \times p}$$

  Consider n-vectors as *one-column* matrices to make all of these definitions consistent. Requiring the dimensions in multiplication to align is a good way to verify linear algebra claims. (*E.g.*, $\boldsymbol{AA}$ does not work for non-square $\boldsymbol{A}$)
- A matrix inverse satisfies $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$. Only full-rank square matrices have inverses
- An (square) orthogonal matrix $\boldsymbol{Q}$ satisfies $\boldsymbol{Q}^\top = \boldsymbol{Q}^{-1}$. If we take the first $n' \leq n$ columns (rows) of an orthogonal matrix we have $\boldsymbol{Q}_{1:n'\cdot}\boldsymbol{Q}_{1:n'\cdot}^\top = \boldsymbol{I}_{n' \times n'}$ so it's transpose-inverse along the 'short-side'
- Any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has a *singular value decomposition*: $\boldsymbol{A} = \boldsymbol{UDV}^\top$ where $r = \min\{m, n\}$, $\boldsymbol{D}$ is a non-negative diagonal $r \times r$ matrix, $\boldsymbol{U} \in \mathbb{R}^{m \times r}$ is the first r columns of an orthogonal $m \times m$-matrix, and $\boldsymbol{V} \in \mathbb{R}^{n \times r}$ is the first r columns of an orthogonal $n \times n$ matrix
- Distributive rules: $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$ and $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ (if all defined)

Matrix Calculus:

- Quadratics: $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \implies \nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}$; $f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = 2\boldsymbol{x}$
- Chain rule: $g(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x}) \implies \nabla g(\boldsymbol{x}) = \boldsymbol{A}^\top (\nabla f)(\boldsymbol{A}\boldsymbol{x})$

Convexity:

- A function $f : \mathbb{R}^p \to \mathbb{R}$ is *convex* if

$$f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y}) \text{ for all } \lambda \in [0, 1], \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$$

  If f is convex and second-differentiable at a point, its second derivative matrix is *positive semi-definite*
- A set $\mathcal{C} \in \mathbb{R}^p$ is convex if

$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C} \implies \lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in \mathcal{C} \text{ for all } \lambda \in [0, 1], \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$$

- If $\nabla f(\boldsymbol{x}_*) = 0$ for convex $f(\cdot)$, then $\boldsymbol{x}_*$ is a global minimizer of $f(\cdot)$

Gradient Methods:

- Given an optimization problem $\min_{\boldsymbol{x} \in \mathcal{C}} f(\boldsymbol{x})$, gradient descent works by repeating the following update:

$$\boldsymbol{x}^{(k+1)} \to \boldsymbol{x}^{(k)} - c\nabla f(\boldsymbol{x}^{(k)})$$

  If $c > 0$ is sufficiently small and $\mathcal{C} = \mathbb{R}^p$, $\boldsymbol{x}^{(k)}$ will converge to a minimizer of f