COMP3670: Introduction to Machine Learning

Release Date. Aug 21st, 2020
Due Date. 23:59pm, Sep 13th, 2020
Maximum credit. 100

Exercise 1 **Orthogonal Compliments** (10+10 credits)

Let V be a vector space, together with an inner product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ and let X and Y be vector subspaces of V. We define the *orthogonal compliment* X^T as

$$X^T := \left\{\mathbf{v} \in V : \langle\mathbf{x},\mathbf{v}\rangle = 0 \text{ for all } \mathbf{x} \in X\right\}$$

1. Prove that $X \cap X^T = \{\mathbf{0}\}$, where $\mathbf{0}$ is the zero vector in V.
2. Prove that if $X \subseteq Y$, then $Y^T \subseteq X^T$.

Exercise 2 **Norms and Inner Products** (10+20 credits)

1. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Let

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) := \frac{\langle\mathbf{v},\mathbf{u}\rangle}{\langle\mathbf{u},\mathbf{u}\rangle}\mathbf{u}$$

   denote the vector projection of $\mathbf{v}$ onto $\mathbf{u}$. Prove that $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ and $\mathbf{u}$ are orthogonal.

2. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Let $||\mathbf{x}|| := \sqrt{\langle\mathbf{x},\mathbf{x}\rangle}$. Prove that $||\cdot||$ is a norm.
   (Hint: To prove the triangle inequality holds, you may need the Cauchy-Schwartz inequality, $\langle\mathbf{x},\mathbf{y}\rangle \leq ||\mathbf{x}||||\mathbf{y}||$.)

Exercise 3 **Vector Calculus** (10+10+30 credits)

1.

$$f, g : \mathbb{R}^n \to \mathbb{R}, \quad f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x}, \quad \boldsymbol{c} \in \mathbb{R}^n, \quad g(\boldsymbol{x}) = \sqrt{\boldsymbol{c}^T\boldsymbol{x} + \mu^2}, \quad \mu \in \mathbb{R}.$$

   - a) (3 points) Prove $\frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = \boldsymbol{c}^T$.
   - b) (2 points) Calculate $\frac{\mathrm{d}g}{\mathrm{d}\boldsymbol{x}}$.

2. Given a system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, with $\boldsymbol{A} \in \mathbb{R}^{k\times n}, \boldsymbol{x} \in \mathbb{R}^{n\times 1}, \boldsymbol{b} \in \mathbb{R}^{k\times 1}$, sometimes there exists no solutions $\boldsymbol{x}$. So we'd like to find a approximate solution $\boldsymbol{A}\boldsymbol{x} \approx \boldsymbol{b}$. To achieve this, we formulate the following regularized least squares error

$$\ell(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2\,, \text{where } \lambda \in$$

   Show that the gradient of the regularized least squares error above is given by

$$\frac{\mathrm{d}\ell(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = 2(\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{b}^T\boldsymbol{A}) + 2\lambda\boldsymbol{x}^T$$

   (Hint: you can directly use the conclusions from questions 2 and 3 above, together with the definition of the Euclidean norm.)