EECS 445: Machine Learning **Fall 2016**

Discussion 1: Linear Algebra

Written by: Chansoo Lee

1.1 Linear Algebra Basics

1.1.1 Dot products

Let $\mathbf{x} = (x_1, \ldots, x_n)^\top, \mathbf{y} = (y_1, \ldots, y_n)^\top$ be column vectors in $\mathbb{R}^n$. The dot product (inner product) of $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

Sometimes we also use the notation $\langle \mathbf{x}, \mathbf{y} \rangle$. We often do not use boldfaced letters for vectors when it's clear from context.

Excercise 1.1. *What is the geometric meaning of dot product?*

1.1.2 Matrix multiplication

In machine learning, it is often convenient to think of a matrix as a vector of vectors. $A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix}$

Let's write matrix-vector multiplication as dot products

$$(A\mathbf{x})_i = \mathbf{a}_i^\top \mathbf{x}.$$

Understand how $A\mathbf{x}$ can be seen as a transformation of $\mathbf{x}$.

Tip: Check dimensions when you multiply matrices and vectors! Typically vectors are column vectors, or equivalently, n-by-1 matrix.

Example 1.1. *Prove* $(AB)^\top = B^\top A^\top$. *It follows* $(A\mathbf{x})^\top = \mathbf{x}^\top A^\top$

Proof: Let $C = (AB)^\top$. Then, $C_{ij} = (AB)_{ji} = \mathbf{a}_j^\top \mathbf{b}_{:i} = \mathbf{b}_{:i}^\top \mathbf{a}_j = (B^T A^T)_{ij}$ ■

Excercise 1.2. *Prove* $(A^\top A)$ *is symmetric.*

1.1.3 Eigenvalues

In this subsection, let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. For all symmetric matrices, we have an eigendecomposition:

$$A = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$$

where $\mathbf{u}_i$'s form an orthonormal basis such that $\|\mathbf{u}_i\|_2 = 1$ and $\mathbf{u}_i^\top \mathbf{u}_j = 0 \ \forall i \neq j$.

Check the following property using the above decomposition:

$$A^k \mathbf{u}_i = \lambda^k \mathbf{u}_i.$$

Excercise 1.3. *Prove that* A *has non-negative eigenvalues if and only if*$\mathbf{x}^\top A \mathbf{x} \geq 0 \ \forall \mathbf{x}$.

Given a vector $\mathbf{x} = \sum \alpha_i \mathbf{u}_i$,

$$A\mathbf{x} = \sum (\alpha_i \lambda_i \mathbf{u}_i^\top \mathbf{x}) \mathbf{u}_i.$$

Understand the geometric meaning of eigenvalues and eigenvectors.

1.2 Multivariable Calculus

We deal with matrices so frequently in machine learning that we will feel tired of elementwise calculations very soon. So, we introduce some conventions and derive useful results. This is important because we want to vectorize our implementations of convex optimization algorithms including gradient descent. Don't be afraid! It's the same multivariable calculus in disguise.

Recall that if $f(\mathbf{x})$ is a scalar, then the we have the gradient vector whose i-th coordinate is:

$$\left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right)_i = \frac{\partial f(\mathbf{x})}{\partial x_i}$$

Example 1.2. *Compute* $\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{x}}$. *Also show* $\frac{\partial \mathbf{y}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{x}}$.

Proof: Let $\mathbf{z}$ be the derivative vector, and $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^n x_i y_i$. Since

$$\mathbf{z}_i = \frac{\partial f}{\partial x_i} = y_i,$$

the derivative is $\mathbf{y}$. (Or it is okay to say the derivative is $\mathbf{y}^\top$. Switching between row and column representation of vector is fine.). ■

Excercise 1.4. *Prove*

$$\frac{\partial \mathbf{y}^\top A\mathbf{x}}{\partial \mathbf{x}} = \mathbf{y}^\top A$$

$$\frac{\partial \mathbf{y}^\top A\mathbf{x}}{\partial \mathbf{y}} = A\mathbf{x}$$

Proof: For the first statement, treat $\mathbf{y}^\top A$ as one vector. ■

1.3 Matrix Calculus

We extend this into vector valued functions (the output can be row or column vector, it's the same):

$$\left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right)_{ij} = \frac{\partial f(\mathbf{x})_i}{\partial x_j}$$

Example 1.3. *Show that*

$$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top}{\partial \mathbf{x}} = I$$

Example 1.4. *Compute* $\frac{\partial A\mathbf{x}}{\partial \mathbf{x}}$.

Proof: Note $(A\mathbf{x})_i = \sum_k a_{ik} x_k$. So, $\frac{\partial (A\mathbf{x})_i}{\partial x_j} = a_{ij}$. ■

Excercise 1.5. *Show*

$$\frac{\partial \mathbf{x}^\top A}{\partial \mathbf{x}} = A^\top$$

1.3.1 Chain Rules

Lemma 1.1. *We have chain rules for vectors* $\mathbf{a}, \mathbf{b}, \mathbf{z}$ *that are functions of* $\mathbf{x}$.

$$\frac{\partial A\mathbf{z}}{\partial \mathbf{x}} = A\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

$$\frac{\partial \mathbf{a}^\top \mathbf{b}}{\partial \mathbf{x}} = \mathbf{a}^\top \frac{\partial \mathbf{b}}{\partial \mathbf{x}} + \mathbf{b}^\top \frac{\partial \mathbf{a}}{\partial \mathbf{x}}$$

Lemma 1.2. $\|\mathbf{x}\|_2^2 = \sum x_i^2 = \mathbf{x}^\top \mathbf{x}$.

Example 1.5. *Compute*

$$\frac{\partial \|\mathbf{x}\|_2^2}{\partial \mathbf{x}}$$

Proof: $\|\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{x}$. Apply Lemma 1.1 with $\mathbf{a} = \mathbf{b} = \mathbf{x}$. ■

Excercise 1.6. *Compute*

$$\frac{\partial \mathbf{x}^\top A\mathbf{x}}{\partial \mathbf{x}}$$

Proof: Apply Lemma 1.1 with $\mathbf{a} = \mathbf{x}$ and $\mathbf{b} = A\mathbf{x}$, and get $\mathbf{x}^\top A + (A\mathbf{x})^\top = \mathbf{x}^\top (A + A^\top)$. ■

Excercise 1.7. *Compute*

$$\frac{\partial \|A\mathbf{x}\|_2^2}{\partial \mathbf{x}}$$

Proof: $\|A\mathbf{x}\|_2^2 = (A\mathbf{x})^\top (A\mathbf{x}) = \mathbf{x}^\top A^\top A\mathbf{x}$. Use the previous result to get $2\mathbf{x}^\top A^T A$ ■

Excercise 1.8. *You'll compute*

$$\frac{\partial \|A\mathbf{x} - \mathbf{y}\|_2^2}{\partial \mathbf{x}}$$

for your homework!