CS 189: Introduction to Machine Learning

Fall 2017

Discussion 1

Due on Friday, August 25th, 2017 at 4 p.m.

Solutions by

Ninh DO

25949105

Problem 1: Unitary Invariance

Prove that the regular Euclidean norm (also called the 2-norm) is unitary invariant; in other words, the 2-norm of a vector is the same, regardless of how you apply a rigid transformation to the vector (i.e., rotate or reflect). Note that rigid transformation of a vector $\mathbf{v} \in \mathbb{R}^d$ means multiplying by an orthogonal $\mathbf{U} \in \mathbb{R}^{d \times d}$.

Let $\mathbf{U} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_n} \end{bmatrix}$ orthogonal, i.e. $\mathbf{u_i} \cdot \mathbf{u_i} = 1$ and $\mathbf{u_i} \cdot \mathbf{u_j} = 0$ for $i \neq j$.

We prove $||\mathbf{a}^T\mathbf{U}|| = ||\mathbf{a}||$ for all $\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T$, or $||\mathbf{a}^T\mathbf{U}||^2 = ||\mathbf{a}||^2$ where $||\cdot||$ is 2-norm.

$$\begin{aligned}
\text{RHS} &= \sum_{i=1}^{n} a_i^2 && (1)\\
\text{LHS} &= \left(\sum_{i=1}^{n} a_i \mathbf{u_i}\right) \cdot \left(\sum_{i=1}^{n} a_i \mathbf{u_i}\right) && (2)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathbf{u_i} \cdot \mathbf{u_j}\\
&= \sum_{i=1}^{n}\sum_{j=1, j=i}^{n} a_i a_j \cancelto{1}{\mathbf{u_i} \cdot \mathbf{u_j}} + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} a_i a_j \cancelto{0}{\mathbf{u_i} \cdot \mathbf{u_j}}\\
&= \sum_{i=1}^{n} a_i^2
\end{aligned}$$

Therefore, LHS = RHS

Problem 2: Eigenvalues

1. Let $\mathbf{A}$ be an invertible matrix. Show that if $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ , then it is also an eigenvector of $\mathbf{A}^{-1}$ with eigenvalue λ^{-1}.

   $\mathbf{A}$ has eigenvalue λ associated with eigenvector $\mathbf{v}$:

$$\begin{aligned}
\mathbf{A}\mathbf{v} &= \lambda\mathbf{v} \\
\mathbf{A}^{-1}\mathbf{A}\mathbf{v} &= \mathbf{A}^{-1}\lambda\mathbf{v} \\
\mathbf{I}\mathbf{v} &= \mathbf{A}^{-1}\lambda\mathbf{v} \\
\lambda^{-1}\mathbf{v} &= \lambda^{-1}\mathbf{A}^{-1}\lambda\mathbf{v} \\
\lambda^{-1}\mathbf{v} &= \mathbf{A}^{-1}\mathbf{v}
\end{aligned}$$

   Therefore, λ^{-1} is the eigenvalue of the matrix $\mathbf{A}^{-1}$ corresponding to the eigenvector $\mathbf{v}$

2. A square and symmetric matrix $\mathbf{A}$ is said to be positive semidefinite (PSD) ($\mathbf{A} \succeq 0$) if $\forall \mathbf{v} \neq \mathbf{0}, \mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0$. Show that $\mathbf{A}$ is PSD if and only if all of its eigenvalues are nonnegative.
   Hint: Use the eigendecomposition of the matrix $\mathbf{A}$.

   $\mathbf{A}$ is symmetric so it can be decomposed into $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where:

   $\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$ is the diagonal matrix of eigenvalues of $\mathbf{A}$ and

   $\mathbf{Q} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_n} \end{bmatrix}$ is the orthogonal matrix formed by the eigenvectors $\mathbf{u_i}$'s of $\mathbf{A}$.

   For any $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^T$, we can write:

$$\begin{aligned}
\mathbf{v}^T\mathbf{A}\mathbf{v} &= \mathbf{v}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{v} \\
&= \begin{bmatrix} \mathbf{v}^T \cdot \mathbf{u_1} & \mathbf{v}^T \cdot \mathbf{u_2} & \cdots & \mathbf{v}^T \cdot \mathbf{u_n} \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u_1}^T \cdot \mathbf{v} \\ \mathbf{u_2}^T \cdot \mathbf{v} \\ \vdots \\ \mathbf{u_n}^T \cdot \mathbf{v} \end{bmatrix} \\
&= \sum_{i=1}^{n} \left(\mathbf{v}^T \cdot \mathbf{u_i}\right) \lambda_i \left(\mathbf{u_i}^T \cdot \mathbf{v}\right) = \sum_{i=1}^{n} \left(\mathbf{v}^T \cdot \mathbf{u_i}\right)^2 \lambda_i
\end{aligned}$$

   If $\mathbf{A}$ has nonnegative eigenvalues, i.e. λ_i's > 0, then $\mathbf{v}^T\mathbf{A}\mathbf{v} = \sum_{i=1}^{n} \left(\mathbf{v}^T \cdot \mathbf{u_i}\right)^2 \lambda_i > 0, \forall \mathbf{v}$, so $\mathbf{A}$ is PSD.

   If $\mathbf{A}$ is PSD, i.e. $\mathbf{v}^T\mathbf{A}\mathbf{v} = \sum_{i=1}^{n} \left(\mathbf{v}^T \cdot \mathbf{u_i}\right)^2 \lambda_i > 0, \forall \mathbf{v}$, assuming that there is at least one eigenvalue is negative, say λ_1, we can choose $\mathbf{v}$ so that $\mathbf{v} \cdot \mathbf{u_1} > 0$ and $\mathbf{v} \cdot \mathbf{u_i} = 0, \forall i \neq 1$. That is to solve the linear equation $\mathbf{v}^T\mathbf{Q} = \mathbf{b}$, where $\mathbf{b} = \begin{bmatrix} c & 0 & \cdots & 0 \end{bmatrix}^T$ with $c \neq 0$. Since $\mathbf{Q}$ is orthogonal and has full rank so this linear equation is always solvable and has unique solution, i.e. we can always find $\mathbf{v}$ such that $\mathbf{v}^T\mathbf{A}\mathbf{v} = \sum_{i=1}^{n} \left(\mathbf{v}^T \cdot \mathbf{u_i}\right)^2 \lambda_i = \left(\mathbf{v}^T \cdot \mathbf{u_1}\right)^2 \lambda_1 < 0$. Contradiction.

Problem 3: Least Squares (using vector calculus)

1. In ordinary least-squares linear regression, there is typically no $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{y}$ (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 $\mathbf{r} = \mathbf{A}\mathbf{x} - \mathbf{y}$ and we want to make it as small as possible. The most common case is to measure the residual error with the standard Euclidean 2-norm. So the problem becomes:

$$\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2 \tag{3}$$

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

We take the derivative of $||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2$ and set it to zero:

$$\begin{aligned}
\frac{\partial ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2}{\partial \mathbf{x}} &= \frac{\partial (\mathbf{A}\mathbf{x} - \mathbf{y})^T (\mathbf{A}\mathbf{x} - \mathbf{y})}{\partial \mathbf{x}} \\
&= \frac{\partial \left(\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{y}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{y} + \mathbf{y}^T\mathbf{y}\right)}{\partial \mathbf{x}} \\
&= \frac{\partial \left(\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}\right)}{\partial \mathbf{x}} - \frac{\partial \left(2\mathbf{x}^T\mathbf{A}^T\mathbf{y}\right)}{\partial \mathbf{x}} + \cancelto{0}{\frac{\partial \left(\mathbf{y}^T\mathbf{y}\right)}{\partial \mathbf{x}}} \\
&= 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{y} = 0 \\
\Rightarrow \quad \mathbf{x} &= \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y}
\end{aligned} \tag{4}$$

2. What should we do if $\mathbf{A}$ is not full rank?

If $\mathbf{A}$ is not full rank, the equation 4 is underdetermined, as some columns of $\mathbf{A}^T\mathbf{A}$ depend on the others. in such a case, we cannot take the inverse of $\mathbf{A}^T\mathbf{A}$ and the equation has infinite solutions.
What we should do is we will take the pseudo-inverse of $\mathbf{A}^T\mathbf{A}$