ELE 535: Machine Learning and Pattern Recognition Homework 2

Zachary Hervieux-Moore

Monday 1st October, 2018

Exercise 1: Let $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m\times n}$. Find the orthogonal projection of A onto $\text{span}(uv^T)$.

Answer: This is equivalent to projecting onto a line and the proof from class is replicated here.

$$\begin{aligned}
&\min_{B\in\mathbb{R}^{m\times n}} \frac{1}{2}\|A-B\|^2 \\
&\text{s.t. } B \in \text{span}(uv^T) \\
\iff& \alpha^* = \arg\min_{\alpha\in\mathbb{R}} \frac{1}{2}\|A-\alpha uv^T\|^2 \\
\iff& \alpha^* = \arg\min_{\alpha\in\mathbb{R}} \frac{1}{2}\|A\|^2 - \alpha\langle A, uv^T\rangle + \frac{1}{2}\|uv^T\|^2
\end{aligned}$$

First order necessary condition and convexity yields a minimum of

$$\alpha^* = \frac{\langle A, uv^T\rangle}{\|uv^T\|^2}$$

Which gives the projection as

$$B = \frac{\langle A, uv^T\rangle}{\|uv^T\|^2}uv^T$$

Notice that the norm could be any one of the valid matrix norms.

Exercise 2: Norm Invariance under Orthogonal Transformations. Show that for any $A \in \mathbb{R}^{m\times n}$, $Q \in \mathcal{O}_m$, $R \in \mathcal{O}_n$, $\|QAR\|_F = \|A\|_F$. Thus the Frovenius norm is invariant under orthogonal transformations. Similarly, show the induced 2-norm of $A \in \mathbb{R}^{m\times n}$ is invariant under orthogonal transformations.

Answer: First, we show invariance for the Frobenius norm.

$$\begin{aligned}
\|QAR\|_F &= \operatorname{trace}(QAR(QAR)^T) \\
&= \operatorname{trace}(QARR^TA^TQ^T) \\
&= \operatorname{trace}(QAA^TQ^T) \\
&= \operatorname{trace}(AA^TQ^TQ) \\
&= \operatorname{trace}(AA^T) \\
&= \|A\|_F
\end{aligned}$$

Now for the induced 2-norm.

$$\begin{aligned}
\|QAR\|_2 &= \sqrt{\langle QAR, QAR\rangle} \\
&= \sqrt{\langle AR, Q^TQAR\rangle} \\
&= \sqrt{\langle AR, AR\rangle} \\
&= \|AR\|_2 \\
&= \max_{\|x\|_2=1} \|ARx\|_2 \\
&= \max_{\|y\|_2=1} \|Ay\|_2 \\
&= \|A\|_2
\end{aligned}$$

Where the second last step of $y = Rx$ is justified because R is orthogonal and so if $\|x\|_2 = 1$ then $\|Rx\|_2 = 1$. As it is invertible, x can always be retrieved with $x = R^Ty$.

Exercise 3: Let A, B be matrices of appropriate size and $x \in \mathbb{R}^n$. Prove that

a) $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$;

b) $\|AB\|_2 \leq \|A\|_2 \|B\|_2$.

Answer:

a) This is direct from the definition.

$$\begin{aligned} \|A\|_2 &= \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_x} \\ \iff \|A\|_2 &\geq \frac{\|Ax\|_2}{\|x\|_2} \quad \text{ if } x \neq 0 \\ \iff \|A\|_2 \|x\|_2 &\geq \|Ax\|_2 \quad \text{ if } x \neq 0 \end{aligned}$$

Now, if $x = 0$, the inequality is trivially satisfied.

b) We simply use part a) twice

$$\begin{aligned} \|AB\|_2 &= \max_{\|x\|_2 = 1} \|ABx\|_2 \\ &\leq \max_{\|x\|_2 = 1} \|A\|_2 \|Bx\|_2 \\ &\leq \max_{\|x\|_2 = 1} \|A\|_2 \|B\|_2 \|x\|_2 \\ &= \|A\|_2 \|B\|_2 \end{aligned}$$

Exercise 4: For $A, B \in \mathbb{R}^{m\times n}$. Show that $\sigma_1(A+B) \leq \sigma_1(A) + \sigma_1(B)$.

Answer: Using the fact that $\sigma(A)_1 = \|A\|_2$. We then apply the definition of the induced 2-norm and use the triangle inequality for vectors

$$\begin{aligned}
\|A+B\|_2 &= \max_{\|x\|_2} \|(A+B)x\|_2 \\
&\leq \max_{\|x\|_2} \|Ax\|_2 + \|Bx\|_2 \\
&\leq \max_{\|x\|_2} \|Ax\|_2 + \max_{\|y\|_2} \|By\|_2 \\
&= \|A\|_2 + \|B\|_2
\end{aligned}$$

Exercise 5: The Moore-Penrose pseudo-inverse. The Moore-Penrose pseudo-inverse of a matrix $A \in \mathbb{R}^{m \times n}$ is the unique matrix $A^+ \in \mathbb{R}^{n \times m}$ satisfying the following four properties:

a) $A(A^+A) = A$

b) $(A^+A)A^+ = A^+$

c) $(A^+A)^T = A^+A$

d) $(AA^+)^T = AA^+$

Let A have compact SVD $A = U\Sigma V^T$. Show that $A^+ = V\Sigma^{-1}U^T$. Give an interpretation of A^+ in terms of $\mathcal{N}(A)$, $\mathcal{N}(A)^\perp$, and $\mathcal{R}(A)$.

Answer: We simply verify the four properties for A^+.

a)

$$A(A^+A) = U\Sigma V^T(V\Sigma^{-1}U^TU\Sigma V^T) = U\Sigma V^T = A$$

b)

$$(A^+A)A^+ = (V\Sigma^{-1}U^TU\Sigma V^T)V\Sigma^{-1}U^T = V\Sigma^{-1}U^T = A^+$$

c)

$$(A^+A)^T = (V\Sigma^{-1}U^TU\Sigma V^T)^T = I^T = I = V\Sigma^{-1}U^TU\Sigma V^T = A^+A$$

d)

$$(AA^+)^T = (U\Sigma V^TV\Sigma^{-1}U^T)^T = I^T = I = U\Sigma V^TV\Sigma^{-1}U^T = AA^+$$

Note that $A^T = V\Sigma U^T$, thus we have that $\mathcal{R}(A^T) = \mathcal{R}(A^+)$. Using the fact that $\mathcal{R}(A^T) = \mathcal{N}(A)^\perp$. So, $\mathcal{R}(A^+) = \mathcal{N}(A)^\perp$. Also, since Σ and Σ^{-1} are diagonal, then we have $\mathcal{N}(A^T) = \mathcal{N}(A^+)$ or $\mathcal{R}(A)^\perp = \mathcal{N}(A^+)$. Thus, we get the interpretation:

$$\mathcal{N}(A) \oplus \mathcal{R}(A^+) = \mathbb{R}^n$$
$$\mathcal{R}(A) \oplus \mathcal{N}(A^+) = \mathbb{R}^m$$