Homework 4

Math 198: Math for Machine Learning

Due Date: March 4
Name:
Student ID:

1 Projections

Let $\mathbf{P} : V \to V$ be a (not necessarily orthogonal) projection operator, i.e. $\mathbf{P}^2 = \mathbf{P}$.

1. Show that all eigenvalues of $\mathbf{P}$ are either 0 or 1.
Let $\mathbf{v_i}$ be an eigenvector of $\mathbf{P}$ with eigenvalue λ_i. Then $\lambda_i \mathbf{v_i} = \mathbf{P} \mathbf{v_i} = \mathbf{P}^2 \mathbf{v_i} = \lambda^2 \mathbf{v_i}$. So $\lambda_i = \lambda_i^2$, and thus the only possible eigenvalues are 0 or 1.

2. Show that $\mathrm{tr}(\mathbf{P}) = \mathrm{rank}(\mathbf{P})$.
Recall that trace and rank are both invariant under similarity. We proved in Note 4 that a projector is similar to a diagonal projector with its eigenvalues on the diagonal. Since all eigenvalues are 0 or 1, and every 0 eigenvalue corresponds to one fewer dimension in the output, we have $\sum_{\lambda_i \neq 0} \lambda_i = \mathrm{rank}(\mathbf{P})$. Therefore, $\mathrm{tr}(\mathbf{P}) = \mathrm{rank}(\mathbf{P})$.

3. Prove that $\mathbf{P}$ is the identity matrix when restricted to its range. That is, for any vector $\mathbf{v} \in \mathrm{range}(\mathbf{P})$, $\mathbf{P}\mathbf{v} = \mathbf{v}$.
We have that $\mathbf{P}^2 = \mathbf{P} = \mathbf{I}\mathbf{P}$, where $\mathbf{I}$ is the identity matrix on the range of $\mathbf{P}$. This implies that the action of $\mathbf{P}$ on its range is the same as the action of $\mathbf{I}$ on its range, and thus $\mathbf{P}$ is the identity matrix when restricted to its range.

4. Prove that $\mathbf{P}$ is either not invertible or its own inverse.
If $\mathbf{P}$ is invertible, then its eigenvalues must all be 1. Therefore, its rank is $\dim(V)$ and its range is V. Since a projection matrix is the identity when restricted to its range, and the domain of this projection matrix is the same as its range, this projection matrix is the identity matrix on V, and therefore is its own inverse.

2 Using the Spectral Theorem

1. Prove that the matrix
$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$
is normal.
Observe that $\mathbf{A}$ is symmetric. Therefore, $\mathbf{A}^\top = \mathbf{A}$, so $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^2 = \mathbf{A}^\top \mathbf{A}$. So, $\mathbf{A}$ is normal. This can also be confirmed by computing $\mathbf{A}\mathbf{A}^\top$ and $\mathbf{A}^\top \mathbf{A}$ and confirming that they are equal.

2. Compute the eigenvalues of $\mathbf{A}$.
Recall that the eigenvalues λ are the solutions to the equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. So, we have $0 = (3-\lambda)(3-\lambda) - 4 = \lambda^2 - 6\lambda + 5$. The roots of this characteristic polynomial are easily found to be $\lambda = 1, 5$.

3. Compute the eigenvectors of $\mathbf{A}$.

For each eigenvalue λ_i, we can find a corresponding eigenvector $\mathbf{v}_i$ by finding a basis for $\ker(\mathbf{A} - \lambda\mathbf{I})$

For $\lambda_1 = 1$, we have

$$\mathbf{A} - \lambda_1\mathbf{I} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

Let $\mathbf{v}_1 = [a_1 \ b_1]^\top$. Then $(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_1 = [2a_1 + b_1 \ 2a_1 + b_1]^\top$ So, to set $\mathbf{v} \in \ker(\mathbf{A} - \lambda_1\mathbf{I})$, we set $a_1 = -b_1$. So $\mathbf{v}_1 = [-1 \ 1]^\top$ is an eigenvector of $\mathbf{A}$. ($\mathbf{A} - \lambda\mathbf{I}$ is dimension 2 and rank 1 by inspection, so the kernel is dimension 1 by Rank-Nullity, and so $\mathbf{v}_1$ is a basis for the kernel.) A similar procedure for $\lambda_2 = 5$ reveals that $\mathbf{v}_2 = [1 \ 1]^\top$ is also an eigenvector of $\mathbf{A}$.

4. Using the results of (a), (b), and (c), combined with your knowledge of the Spectral Theorem, compute the eigendecomposition of $\mathbf{A}$ (in particular, find the orthogonal matrix $\mathbf{Q}$ and diagonal matrix $\mathbf{\Lambda}$ such that $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$).

By the Spectral Theorem for normal matrices, we have that $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$, where

$$\mathbf{Q} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

Note that the columns of $\mathbf{Q}$ must be *orthonormal* vectors. This is why the eigenvectors from (3) have all been divided by $\sqrt{2}$.

5. Compute $\mathbf{A}^{20}$.

Note that $\mathbf{A}^{20} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top\ldots\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top = \mathbf{Q}\mathbf{\Lambda}^{20}\mathbf{Q}^\top$, as $\mathbf{Q}^\top\mathbf{Q} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{I}$. Additionally, we have

$$\mathbf{\Lambda}^{20} = \begin{bmatrix} 1^{20} & 0 \\ 0 & 5^{20} \end{bmatrix}$$

as $\mathbf{\Lambda}$ is diagonal. So,

$$\mathbf{A}^{20} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5^{20} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{5^{20}}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{5^{20}}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1+5^{20}}{2} & \frac{-1+5^{20}}{2} \\ \frac{-1+5^{20}}{2} & \frac{1+5^{20}}{2} \end{bmatrix}$$

3 PSD Matrices

1. Prove that the two definitions given for positive semi-definite matrices are equivalent. *(Hint: use Rayleigh quotients.)*

Suppose $\mathbf{A}$ is a symmetric matrix such that for any vector $\mathbf{v}$, $\mathbf{v}^\top\mathbf{A}\mathbf{v} \geq 0$. Let $\mathbf{v_i}$ be an eigenvector of $\mathbf{A}$. Then $0 \leq \mathbf{v_i}^\top\mathbf{A}\mathbf{v} = \lambda_i\mathbf{v}^\top\mathbf{v} = \lambda_i||\mathbf{v}||^2$. Since $||v||^2 > 0$, this implies $\lambda_i \geq 0$, and so all eigenvalues must be non-negative.

Suppose $\mathbf{A}$ is a symmetric matrix with all non-negative eigenvalues. Then for all non-zero $\mathbf{v}$, $0 \leq \lambda_{\min}(\mathbf{A}) \leq R_\mathbf{A}(\mathbf{v})$, where $R_\mathbf{A}(\cdot)$ denotes the Rayleigh quotient. Since $\mathbf{v}^\top\mathbf{A}\mathbf{v}$ has the same sign as $R_\mathbf{A}(\mathbf{v})$, we thus have that $\mathbf{v}^\top\mathbf{A}\mathbf{v} \geq 0$ for all $\mathbf{v}$.

2. Prove that, for any matrix $\mathbf{X}$ and any scalar $\lambda > 0$, $\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}$ is invertible.

Observe that, for all nonzero $\mathbf{v}$, $\mathbf{v}^\top(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})\mathbf{v} = \mathbf{v}^\top\mathbf{X}^\top\mathbf{X}\mathbf{v} + \lambda\mathbf{v}^\top\mathbf{I}\mathbf{v} = ||\mathbf{X}\mathbf{v}||^2 + \lambda||\mathbf{v}||^2 > 0$ (since $||\mathbf{v}|| \neq 0$). So, $\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}$ is positive definite and thus invertible.