# Homework 2 Solutions

Math 198: Math for Machine Learning

Due Date: February 19
Name: Sean Vernon

## 1 Comparing Vector Spaces

1. Exhibit a basis for $\mathbb{R}^3 := \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$.
   One basis for $\mathbb{R}^3$ is the standard basis, $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. Any three linearly independent vectors in $\mathbb{R}^3$ will do.

2. Exhibit a basis for $\mathbb{P}^2 := \{f : \mathbb{R} \to \mathbb{R} : f(x) = a_0 + a_1x + a_2x^2 \text{ for some } a_0, a_1, a_2 \in \mathbb{R}\}$, the space of 3rd degree polynomials with real coefficients. Note that your basis elements should be polynomials.
   The standard basis for $\mathbb{P}^2$ is $\{1, x, x^2\}$.

3. Conclude that $\mathbb{R}^3$ and $\mathbb{P}^2$ are *isomorphic* (i) by a dimension argument and (ii) by exhibiting an *isomorphism* between them. When two vector spaces $V, W$ are isomorphic, we write $V \cong W$.
   As noted in class, any vector space with dimension $d$ is isomorphic to $\mathbb{R}^d$. $\mathbb{P}^2$ has dimension 3, as there are 3 elements in the basis. Therefore, $\mathbb{P}^2 \cong \mathbb{R}^3$. We can also exhibit the following isomorphism between them. Let $e_i$ denote the $i$-th element of the standard basis for $\mathbb{R}^3$ and $f_i$ denote the $i$-th element of the standard basis for $\mathbb{P}^2$. Then $T : \mathbb{R}^3 \to \mathbb{P}^2$ given by $e_i \mapsto f_i$ is an isomorphism, as its inverse $T^{-1} : \mathbb{P}^2 \to \mathbb{R}^3$ given by $f_i \mapsto e_i$ exists. So, $\mathbb{R}^3 \cong \mathbb{P}^2$.

## 2 Characterizing the Inner Product

1. Let $\langle \cdot, \cdot \rangle$ be an inner product on $\mathbb{R}^n$. Show that there exists $A \in \mathbb{R}^{n \times n}$ with $A^\top = A$ such that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top A\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. (Hint: what is the action of $\langle \cdot, \cdot \rangle$ on the standard basis?)
   Define a matrix $A$ such that $A_{ij} = \langle e_i, e_j \rangle$. Then for any two vectors $v, w \in \mathbb{R}^n$, we have

$$\langle v, w \rangle = \langle \sum_{i=1}^{n} \alpha_i e_i, \sum_{j=1}^{n} \beta_j e_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \alpha_i e_i, \beta_j e_j \rangle$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j \langle e_i, e_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j A_{ij}$$

$$= v(\sum_{j=1}^{n} \beta_j A_j) = vAw^\top$$

   From here, it suffices to show that $A^\top = A$. By definition, we have that $A_{ij} = \langle e_i, e_j \rangle = \langle e_j, e_i \rangle = A_{ji}$. This completes the proof.

## 3 Linear Maps

Let $V, W$ be vector spaces, and let $T : V \to W$ be a linear map.

1. Show that $T$ is one-to-one (a.k.a. injective) if and only if the kernel of $T$ is trivial, i.e. $\{v \in V : T(v) = \mathbf{0}_W\} = \{\mathbf{0}_V\}$.
   ($\rightarrow$) If $T$ is one-to-one, then no two unique vectors $v_1, v_2 \in V$ map to the same vector $w \in W$. Therefore, only one vector in $V$ maps to $0_W$. Since $T(0_V) = 0_W$, $0_V$ is the only vector in $V$ which maps to $0_W$, and so $\ker(T) = \{0_v\}$ (i.e. the kernel is trivial). $\square$
   ($\leftarrow$) Let the kernel of $T$ be trivial. Suppose there are two unique vectors $v_1, v_2 \in V$ such that $T(v_1) = T(v_2)$. But then $T(v_1 - v_2) = T(v_1) - T(v_2) = 0$, which implies $v_1 - v_2 \in \ker(T)$. Since the kernel of $T$ is trivial, this in turn implies $v_1 - v_2 = 0$, and so $v_1 = v_2$. This contradicts our assumption that $v_1$ and $v_2$ are not the same vector, and so no two unique vectors in $V$ map to the same vector in $W$. Therefore, $T$ is one-to-one. $\square$

2. Let $\{\mathbf{b}_1, ..., \mathbf{b}_n\}$ be a basis for $V$, and let $T$ be such that $\{T(\mathbf{b}_1), ..., T(\mathbf{b}_n)\}$ is a basis for $W$. Show that $T$ is an isomorphism.
   We first show that $T$ is onto – since $\{T(\mathbf{b}_1), \ldots, T(\mathbf{b}_n)\}$ is a basis for $W$, the range of $T$ spans $W$. Furthermore, since no two basis vectors in $V$ map to the same basis vector in $W$ under $T$, no two vectors in $V$ will map to the same vector in $W$, and so $T$ is one-to-one. Therefore, $T$ is an isomorphism.

# 4 Dual Spaces (Optional)

Given a vector space $V$, we can form the vector space of linear maps from $V$ to $\mathbb{R}$, called the *dual space* of $V$. Formally, the dual space is given by $V^* := L(V, \mathbb{R})$, and its elements are known as *linear functionals* (or, in some contexts, *covectors*). *Note that the questions in this section are optional.*

1. Let $V$ be a real vector space with basis $B = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$. Exhibit a basis for $V^*$ and conclude that $V \cong V^*$.
   Recall that any linear map is defined by its action on a basis. So, we can exhibit a basis for $V^*$ by considering the set of linear maps which map each individual basis vector in $V$ to $1 \in \mathbb{R}$. Then, by scaling and combining these linear maps, we can create linear maps with arbitrary actions on the basis of $V$, and thus this set will span all linear maps in $V^*$ while remaining linearly independent, our condition for a basis. This basis is $\{\delta_1, \delta_2, \ldots, \delta_n\}$ where $\delta_i$ is the function which maps $b_i$ to 1 and all other basis vectors to 0. Since this basis has $n$ elements, $\dim(V^*) = n$, and so $V \cong V^*$.

2. Consider $V^{**}$, the dual space of the dual space of $V$, called the *double dual space* of $V$. Without choosing a basis for $V$, construct an isomorphism between $V$ and $V^{**}$. Since such an isomorphism exists, we say that $V$ and $V^{**}$ are *canonically isomorphic*.
   Since $V^*$ is a space of functions, and we wish to map those functions to the real numbers, one can consider the elements of $V^{**}$ as being the "arguments" to those functions, i.e. the $i$-th basis element of $V^{**}$ would be the vector which maps $\delta_i$ to 1 and all other basis elements of $V^*$ to 0.

   More generally, define a map $\Phi : V \to V^{**}$ given by $x \mapsto \phi_x$, where $\phi_x$ is the evaluation map given by $\phi_x(\xi) = \xi(x)$. By the linearity of elements of $V^*$ (the $\xi$'s), $\Phi$ is linear. We have already established that $V \cong V^{**}$, and $\Phi$ is clearly one-to-one. Thus, $\Phi$ is the canonical isomorphism we're looking for.

3. Let $H$ be a (real) Hilbert space, i.e. an inner product space (perhaps infinite-dimensional) that is complete with respect to the metric induced by its inner product. Form the *continuous dual space* of $H$, $H' = \{\xi \in H^* : \xi \text{ is continuous}\}$. It turns out that $H$ is canonically isomorphic to its continuous dual; this result in functional analysis known as the Riesz Representation Theorem. Give a guess as to the canonical isomorphism $H \to H'$. (Hint: It depends completely upon the coordinate-wise linearity of the inner product.)
   Define $\Psi : V \to V'$ given by $\Psi(x)(y) = \langle y, x \rangle$. Our $\Psi$ maps $x$ to the functional $y \mapsto \langle y, x \rangle$. If we call this functional $\psi_x$, then $||\psi_x||_{\text{op}} = ||x||_H$ so $\psi_x$ is bounded and thus continuous. Since $\mathbb{R}^n$ is a Hilbert space with the standard inner product, every linear functional is of the form $\langle \cdot, x \rangle$ for some $x \in \mathbb{R}^n$.

## 5 Projections

A set of vectors is *orthogonal* if each vector is pairwise orthogonal to all the rest. An orthogonal set of vectors is *orthonormal* if each vector has norm 1 (i.e., $\langle \mathbf{v}, \mathbf{v} \rangle = 1$).

1. Show that any set of orthonormal vectors is linearly independent.
   We first show that only the zero vector is orthogonal to itself, as in the definition of inner product, $\langle v, v \rangle = 0$ if and only if $v = 0$. Note that the zero vector could never be a member of a set of orthonormal vectors, as its norm is not 1 (it is 0). Suppose we have a set of orthonormal vectors $\{v_1, \ldots, v_n\}$ which is not linearly independent. Then some vector in this set can be written as a linear combination of the others. Without loss of generality, let this vector be $v_i$. Then

   $$\sum_{j \neq i}^{n} \alpha_j v_j = v_i$$

   However, we have that

   $$\langle \sum_{j \neq i}^{n} \alpha_j v_j, v_i \rangle = \sum_{j \neq i}^{n} \alpha_j \langle v_j, v_i \rangle = 0$$

   since the set is orthonormal. This is a contradiction, as no non-zero vector is orthogonal to itself, and 0 cannot be a member of our set. Therefore, no linearly dependent set of orthonormal vectors exists.

2. Show that, for a space $V$ with dimension $d$, a set of $d$ orthonormal vectors in $V$ is a basis for $V$.
   This follows easily from the previous problem, as any $d$ linearly independent vectors in a space with dimension $d$ constitute a basis. Since $d$ orthonormal vectors are always linearly independent, any set of $d$ orthonormal vectors in $V$ is a basis for $V$.

3. Suppose we have a vector space $V$, a subspace $W \subset V$, and a vector $\mathbf{v} \in V$ such that $\mathbf{v} \notin W$. Let $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be an orthonormal basis for $W$.

   (a) Show that

   $$\mathbf{v}_w = \sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{w}_i \rangle \mathbf{w}_i$$

   is an orthogonal projection of $\mathbf{v}$ into $W$.
   Note that we can extend the orthonormal basis for $W$ to an orthonormal basis for $V$ by adding vectors orthogonal to all those in our current basis until we reach $\dim(V)$ vectors. Suppose we have done so and obtained a basis $\{w_1, \ldots, w_n\}$. Then we can write $v$ in terms of this basis as

   $$v = \alpha_1 w_1 + \ldots + \alpha_n w_n$$

   Note that $\langle v, w_i \rangle = \alpha_i$ as our basis is orthonormal. So,

   $$v_w = \sum_{i=1}^{k} \alpha_i w_i$$

   and

   $$v - v_w = \sum_{i=k+1}^{n} \alpha_i w_i$$

   which is a linear combination of vectors orthogonal to every vector in $W$ and is thus orthogonal to $W$. By the proof in the notes, this implies $v_w$ is an orthogonal projection of $v$ into $W$.

   (b) Let $\mathbf{P}_W$ be a matrix such that $\mathbf{P}_W \mathbf{v} = \mathbf{v}_w$. $\mathbf{P}_W$ is known as a *projection matrix.*

   i. Show that $\mathbf{P}_W^2 = \mathbf{P}_W$.
   Note that $P_W(P_W v)$ will return a vector $v_w$ such that $P_W v - v_w$ is orthogonal to $W$. But both $P_W v$ and $v_w$ are in $W$, which is closed, so $P_W v - v_w \in W$. Therefore, in order for $P_W v - v_w$ to be orthogonal to all $w \in W$, it must be orthogonal to itself, and thus it is the zero vector. So $P_W v - P_W(P_W v) = 0$, and thus $P_W^2 = P_W$.

ii. Show that $\mathbf{P}_W^\top = \mathbf{P}_W$.

Recall that $P_W$ is the matrix such that $P_W v = \sum_{i=1}^{k} \langle v, w_i \rangle w_i$. We can rewrite the expression on the right to get a new matrix form for $P_W$:

$$\sum_{i=1}^{k} \langle v, w_i \rangle w_i = \sum_{i=1}^{k} v^\top w_i w_i = \sum_{i=1}^{k} u_i u_i^\top v = (\sum_{i=1}^{k} u_i u_i^\top) v = UU^\top v$$

where $U$ is a matrix with $u_1, \ldots, u_k$ as its columns. (This last step is known as the *sum-of-outer-products identity.*) Therefore, $P_W = UU^\top = (UU^\top)^\top = P_W^\top$.

iii. Show that $\mathbf{P}_W \mathbf{w} = \mathbf{w}$ for all $w \in W$. Conclude that $\mathbf{P}_W$ has the same action as the identity matrix $\mathbf{I}$ on vectors in $W$.

This proof follows from the proof in part i. Note that since any vector in $W$ has the first same $k$ components as an infinite number of vectors in $V$, $\text{range}(P_W) = W$. Since $P_W^2 = P_W$, $P_W^2 v = IP_W v$. This implies that since $P_W v$ is always a vector in $W$, the action of $P_W$ on a vector in $W$ is the same as the action of $I$ on a vector in $W$.

4. Let $V = \mathbb{R}^4$ and $U \cong \mathbb{R}^2$ such that $\{e_1, e_2\}$ is a basis for $U$ (i.e. all vectors in $U$ take the form $(a, b, 0, 0)$ with respect to the standard basis for $\mathbb{R}^4$). Note that $U$ is a subspace of $V$. Determine the matrix form of $\mathbf{P}_U$, and show that it has the same properties as $\mathbf{P}_W$ from problem 3.

As noted in 3a, $\sum_{i=1}^{k} \langle v, u_i \rangle u_i$ is an orthogonal projection of $v$ into $U$. Using the bases given in this problem, we can determine the action of this summation on the standard basis for $\mathbb{R}^4$, and use this to define a matrix $P_U$. We have that:

$$\sum_{i=1}^{2} \langle e_1, e_i \rangle e_i = e_1, \; \sum_{i=1}^{2} \langle e_2, e_i \rangle e_i = e_2$$

$$\sum_{i=1}^{2} \langle e_3, e_i \rangle e_i = 0, \; \sum_{i=1}^{2} \langle e_4, e_i \rangle e_i = 0$$

therefore,

$$P_U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Note that

$$P_U^\top = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \; P_U^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ and } P_U u = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ 0 \\ 0 \end{bmatrix} = u$$

as desired.