Topics of the week

1. Define Projection, Derive formula for Projection on a subspace, Compute the Projection Matrix.
2. Show that when A has independent columns, $A^\top A$ is invertible and symmetric.
3. Define Least Squares solution, derive Normal equations, compute a least squares solution.
4. Use Least Squares to fit a line to points (linear regression).

Inner product space is a vector space with well-defined scalar product:

1. Symmetry: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$;
2. Linearity: $(a\mathbf{x} + b\mathbf{y}) \cdot \mathbf{z} = a(\mathbf{x} \cdot \mathbf{z}) + b(\mathbf{y} \cdot \mathbf{z})$;
3. Positive-definite: $\mathbf{x} \cdot \mathbf{x} \geq 0$ and $\mathbf{x} \cdot \mathbf{x} = 0 \iff \mathbf{x} = 0$.

Note: Not all vector spaces can be equipped with an inner product!
Explanation: Abstractly, a vector space is defined over a field $\mathbb{F}$, and a lot of fields can't be ordered.

Conjugate transform of $A : V \to W$ is $A^\top : W \to V$ such that $Ax \cdot y = x \cdot A^\top y$.
In matrices: $A^\top$ corresponds to the transposed matrix.

Orthogonality Subspaces V and W are orthogonal if $v^\top w = 0$ for all $v \in V$ and $w \in W$.
Criterion: $V \perp W \iff e_i^\top g_j = 0$, where $\{e_i\}$ is the basis of V and $\{g_j\}$ is the basis of W.

Subspace intersection of subspaces V and W is $V \cap W = \{v : v \in V, v \in W\}$.
Orthogonal spaces: $V \perp W \implies V \cap W = \{0\}$, because $v \in V \cap W \implies v \cdot v = 0$.

Subspace sum of subspaces V and W is $V + W = \{v + w : v \in V, w \in W\} = \operatorname{span}(V \cup W)$.

Dimension property $\dim(V + W) + \dim(V \cap W) = \dim V + \dim W$.
Explanation: Take bases of V and W, and reform them into bases of $V + W$ and $V \cap W$.
Orthogonal spaces: $\dim(V + W) = \dim V + \dim W$.

Orthogonal complement of V is $V^\perp = \{w : w^\top v = 0, \forall v \in V\}$.
Interpretation: Union of all orthogonal subspaces of V.
In matrices: $N(A) = R(A)^\top \iff \ker A \perp \operatorname{im} A^\top$, i.e. $Ax = 0 \iff x \cdot A^\top y = Ax \cdot y = 0$ for all y.

Implicit and explicit form Any subspace U can be defined in either of the following ways:

- Explicit form: $U = \{y = Ax : x \in V\} = \operatorname{im} A$, defining y via parametric equation.
- Implicit form: $U = \{y \in V : By = 0\} = \ker B$, defining y via implicit equation.

The basis of U defines A, while the basis of $U^\perp$ defines $B^\top$.

- $(V \cap W)^\perp = V^\perp + W^\perp$, union of implicit equations $\implies$ intersection of spaces they define;
- $(V + W)^\perp = V^\perp \cap W^\perp$, intersection of implicit equations $\implies$ span of the union of defined spaces.

Connection between A and $A^\top A$ $Ax = 0 \iff \|Ax\| = 0 \iff x^\top (A^\top A)x = 0$.

Orthogonal projection of a vector v on the subspace S is $s \in S$ s.t. $\|v - s\| \to \min$.

Finding projection: Let $\operatorname{im} A = S$ and $\operatorname{im} B = S^\perp$, then $v = Ax + By$ and $s = Az$, so

$$\begin{gathered}
\|v - Az\|^2 \to \min \\
\|v\|^2 - 2v \cdot Az + \|Az\|^2 \to \min \\
(Ax + By) \cdot Az + \|Az\|^2 \to \min \\
Ax \cdot Az + \|Az\|^2 \to \min
\end{gathered}$$

Thus, z only depends on x, and for $v = Ax$, the answer is clearly $z = x$, as $\|Ax - Ax\| = 0$. We can find x from $A^\top v = A^\top Ax + A^\top By = A^\top Ax$, hence $x = (A^\top A)^{-1} A^\top v$.

Least squares Given n observations (a_i, b_i), where $b_i \in \mathbb{R}$ and $a_i \in \mathbb{R}^m$, predict $b_i = f(a_i)$.

Linear model: $b_i \approx x^\top a_i \implies b \approx Ax \implies \|b - Ax\| \to \min$.

Fitting a line: Given (x_i, y_i) with $x, y \in \mathbb{R}$, and the model is $y_i = ax_i + b$:

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \approx \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \iff \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \approx \begin{bmatrix} y \end{bmatrix}$$

Using $x = (A^\top A)^{-1} A^\top$, we get:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} x^\top x & 1^\top x \\ x^\top 1 & 1^\top 1 \end{bmatrix}^{-1} \begin{bmatrix} x^\top y \\ 1^\top y \end{bmatrix} \iff \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i x_i^2 & \sum_i x_i \\ \sum_i x_i & n \end{bmatrix}^{-1} \begin{bmatrix} \sum_i x_i y_i \\ \sum_i y_i \end{bmatrix}$$

Note: If $1^\top x = 0$, then $a = \frac{x^\top y}{x^\top x}$ and $b = \frac{1^\top y}{m}$.