PCA

Basics: vectors have multiple representations in different bases.

$$\begin{bmatrix}1\\2\\3\end{bmatrix}=c_1\begin{bmatrix}1\\0\\0\end{bmatrix}+c_2\begin{bmatrix}0\\1\\0\end{bmatrix}+c_3\begin{bmatrix}0\\0\\1\end{bmatrix}=x_1\begin{bmatrix}1\\2\\0\end{bmatrix}+x_2\begin{bmatrix}0\\0\\2\end{bmatrix}$$

PCA represents the data as a linear combination of principal components.

$$\begin{bmatrix}1\\2\\3\end{bmatrix}=c_1[PC_1]+c_2[PC_2]+...+c_n[PC_n]$$

Rewrite the first system:

$$\begin{bmatrix}1\\2\\3\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}\leftrightarrow v=Bc$$

Projection

The first PC is vector which maximizes the variance of the projection of the data, D, onto that vector.
Def. orthogonal: **a** and **b** are orthogonal if their inner product is 0.

The projection of **a** and **b** is thus given by:

$$proj_b(a)=\frac{a^T b}{b^T b}$$

So for many points (lines, because all points will form a line with the origin) $a_1 ... a_n$, with their respective projections on $\boldsymbol{b}$, $c_1 ... c_n$, we wish to "minimize their variance". Note that we can write variance as:

$$Var(X)=\frac{1}{n}X^T X$$

Now we can properly state our objective:

$$\max_{\|v\|=1} var(proj_v D)$$

The projection of $\boldsymbol{D}$ on $\boldsymbol{v}$ can be written as:

$$\begin{bmatrix}proj_v d_1\\\vdots\\proj_v d_n\end{bmatrix}=\begin{bmatrix}v^T d_1\\\vdots\\v^T d_n\end{bmatrix}=D^T v$$

And we can rewrite $var(\boldsymbol{D^T v})$ in a quadratic form, because quadratic forms are easy to maximize analytically.

$$(D^T v)^T (D^T v)=v^T D D^T v$$

To maximize quadratic forms in general, given:

$$x^T A x, A^T=A, \|x\|=1$$

We use the following theorem about symmetric matrices, stated without proof, if we have a symmetric matrix $\boldsymbol{A}$, then we can decompose it in terms of 2 matrices $\boldsymbol{E}$ and $\boldsymbol{V}$, as follows:

$$A^T=A\Rightarrow A=EVE^T, E^T=E^{-1}, V=\begin{bmatrix}v_1&0\\0&v_2\end{bmatrix}$$

The matrix $\boldsymbol{E}$ is composed of the eigenvectors of $\boldsymbol{A}$, and $\boldsymbol{V}$ are the eigenvalues.
Since $\mathbf{DD^T}$ is necessarily symmetric, we can use that decomposition trick:

$$DD^T=EVE^T$$

Where $\boldsymbol{E}$ is a matrix of the principal components of $\boldsymbol{DD^T}$.

~Fine~