10-601 Recitation

William Bishop
November 9th, 2011

Agenda

- Linear Algebra Review
- Matrix Derivatives
- SVD
- PCA

Linear Algebra Review

$$\begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$$

What does it mean to take the derivative with respect to a vector or a matrix?

Reference Reading: Bishop Appendix C.

Linear Algebra Review

What does it mean to take the derivative with respect to a vector or a matrix?

$$y = \vec{x}^T A \vec{x}$$

$$\frac{\partial y}{\partial \vec{x}} = ?$$

Just do this component wise.

Linear Algebra Review

What does it mean to take the derivative with respect to a vector or a matrix?

$$
\begin{aligned}
f(\vec{x}) &= \vec{x}^T A \vec{x} \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_1x_2 + a_{22}x_2^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= 2x_1a_{11} + (a_{12} + a_{21})x_2 \\
\frac{\partial f}{\partial x_2} &= 2x_2a_{22} + (a_{12} + a_{21})x_1
\end{aligned}
$$

Linear Algebra Review

What does it mean to take the derivative with respect to a vector or a matrix?

$$\frac{\partial f}{\partial x_1} = 2x_1 a_{11} + (a_{12} + a_{21})x_2$$
$$\frac{\partial f}{\partial x_2} = 2x_2 a_{22} + (a_{12} + a_{21})x_1$$

$$\frac{\partial f}{\partial \vec{x}} = \left[\begin{array}{cc} 2a_{11} & a_{12} + a_{21} \\ a_{12} + a_{21} & 2a_{22} \end{array} \right] \vec{x}$$

Linear Algebra Review

What does it mean to take the derivative with respect to a vector or a matrix?

$$\frac{\partial f}{\partial \vec{x}} = \left[\begin{array}{cc} 2a_{11} & a_{12} + a_{21} \\ a_{12} + a_{21} & 2a_{22} \end{array} \right] \vec{x}$$

If A is symmetric:

$$\frac{\partial f}{\partial \vec{x}} = 2A\vec{x}$$

Can work these kind of identities out for matrix/vector equations as well.

Eigen-WHAT?????

Square Matrix.

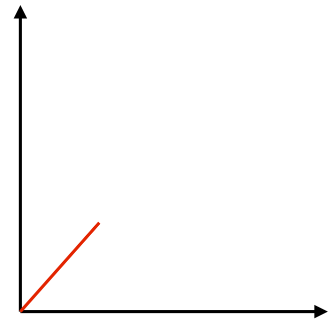
$$\lambda \vec{v} = A \vec{v}$$

Eigenvector

Eigenvalue.

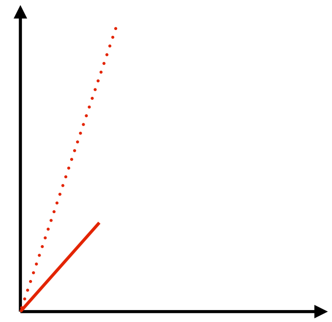
Eigen-WHAT?????

$$\lambda \vec{v} = A \vec{v}$$



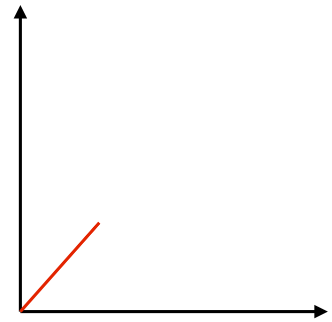
Eigen-WHAT?????

$$\lambda \vec{v} = A \vec{v}$$



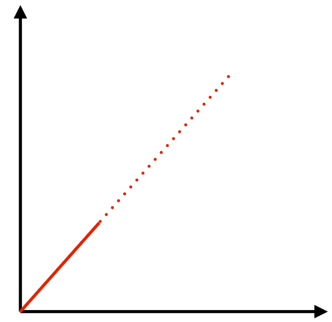
Eigen-WHAT?????

$$\lambda \vec{v} = A \vec{v}$$



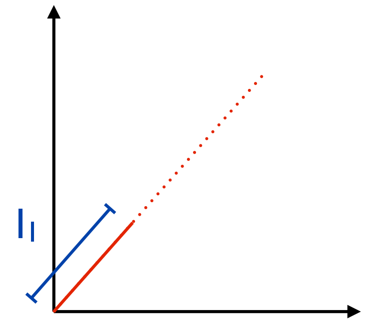
Eigen-WHAT?????

$$\lambda \vec{v} = A\vec{v}$$



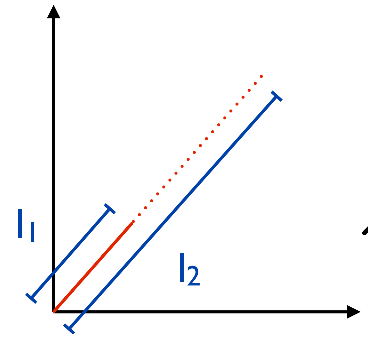
Eigen-WHAT?????

$$\lambda \vec{v} = A\vec{v}$$



Eigen-WHAT?????

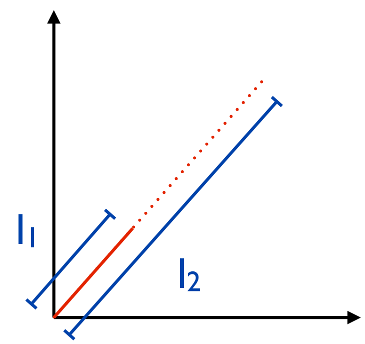
$$\lambda\vec{v} = A\vec{v}$$



$$\lambda = \frac{l_2}{l_1}$$

Eigen-WHAT?????

$$\lambda \vec{v} = A\vec{v}$$



Eigen-WHAT?????

- How do we find eigenvalues?

$$\lambda \vec{v} = A\vec{v}$$
$$\lambda I \vec{v} = A\vec{v}$$
$$0 = [A - \lambda I]\vec{v}$$

DETAILS

Step One: Find eigenvalues for which:

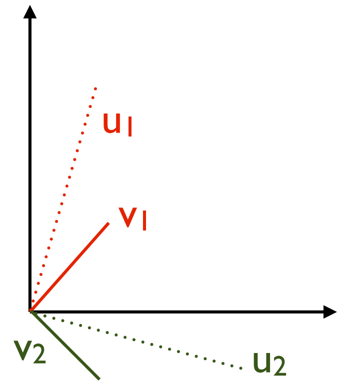
$$\det\left([A - \lambda I]\right) = 0$$

Step Two: Use eigenvalues and solve for eigenvectors:

$$0 = [A - \lambda I]\vec{v}$$

Singular Value Decomposition

Very cool fact: For any matrix, A, I can find a set of orthogonal vectors $\vec{v}_1, \ldots, \vec{v}_n$ such that $\vec{u}_1 = A\vec{v_1}, \ldots, \vec{u}_2 = A\vec{v_2}$ are orthogonal as well.



Singular Value Decomposition

Very cool fact: For any matrix, A, I can find a set of orthogonal vectors $\vec{v}_1, \ldots, \vec{v}_n$ such that $\vec{u}_1 = A\vec{v_1}, \ldots, \vec{u}_2 = A\vec{v_2}$ are orthogonal as well.

- Let $\vec{v_1}$ and $\vec{v_2}$ be orthogonal eigenvectors of $A^T A$ with eigenvalues λ_1, λ_2.
- By definition we have: $\vec{u}_1 = A\vec{v}_1$ and $\vec{u}_2 = A\vec{v}_2$.

Because this is symmetric
we can always find an orthornormal
eigenbasis for it.

$$
\begin{aligned}
\vec{u}_1^T \vec{u}_2 &= (A\vec{v_1})^T A\vec{v}_2 \\
&= \vec{v_1}^T A^T A\vec{v_2} \\
&= \vec{v_1}^T \lambda_2 \vec{v_2} \\
&= 0
\end{aligned}
$$

Singular Value Decomposition

This is great because if we now have an orthonormal basis, $\vec{v}_1 \dots \vec{v}_n$, for A and we know that $\vec{v}_1 \to \sqrt{\lambda_1}\vec{u_1}, \dots \vec{v}_n \to \sqrt{\lambda_n}\vec{u}_n$, so we can write:

$$A = U\Sigma V^T$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \\ a_{51} & a_{52} & a_{53} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}^T$$

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix}$$

Singular Value Decomposition

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \\ u_{41} & u_{42} & u_{43} \\ u_{51} & u_{52} & u_{53} \end{bmatrix}$$

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

Unit vector. Each column is a sample.

$$\vec{v}^T X = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}$$

$$= \begin{bmatrix} \vec{v}^T \vec{x}_1 & \vec{v}^T \vec{x}_2 & \vec{v}^T \vec{x}_3 & \vec{v}^T \vec{x}_4 \end{bmatrix}$$

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

$$\text{Variance of Projected Data} = \vec{v}^T X X^T \vec{v}$$

Want to maximize this subject to $\vec{v}$ being a unit vector.

$$L(\vec{v}) = \left(\vec{v}^T X X^T \vec{v} - \lambda(\vec{v}^T \vec{v} - 1)\right)$$

Lagrange multipliers: next week!

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

$$L(\vec{v}) = \left(\vec{v}^T X X^T \vec{v} - \lambda(\vec{v}^T \vec{v} - 1)\right)$$

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

$$L(\vec{v}) = \left(\vec{v}^T X X^T \vec{v} - \lambda(\vec{v}^T \vec{v} - 1)\right)$$

$$\frac{\partial L}{\partial \vec{v}} = 2XX^T\vec{v} - 2\lambda\vec{v}$$

$$0 = (XX^T - \lambda I)\vec{v}$$

Thus, we are just looking for the eigenvector of this matrix.

Can also think of the equivalent SVD problem.