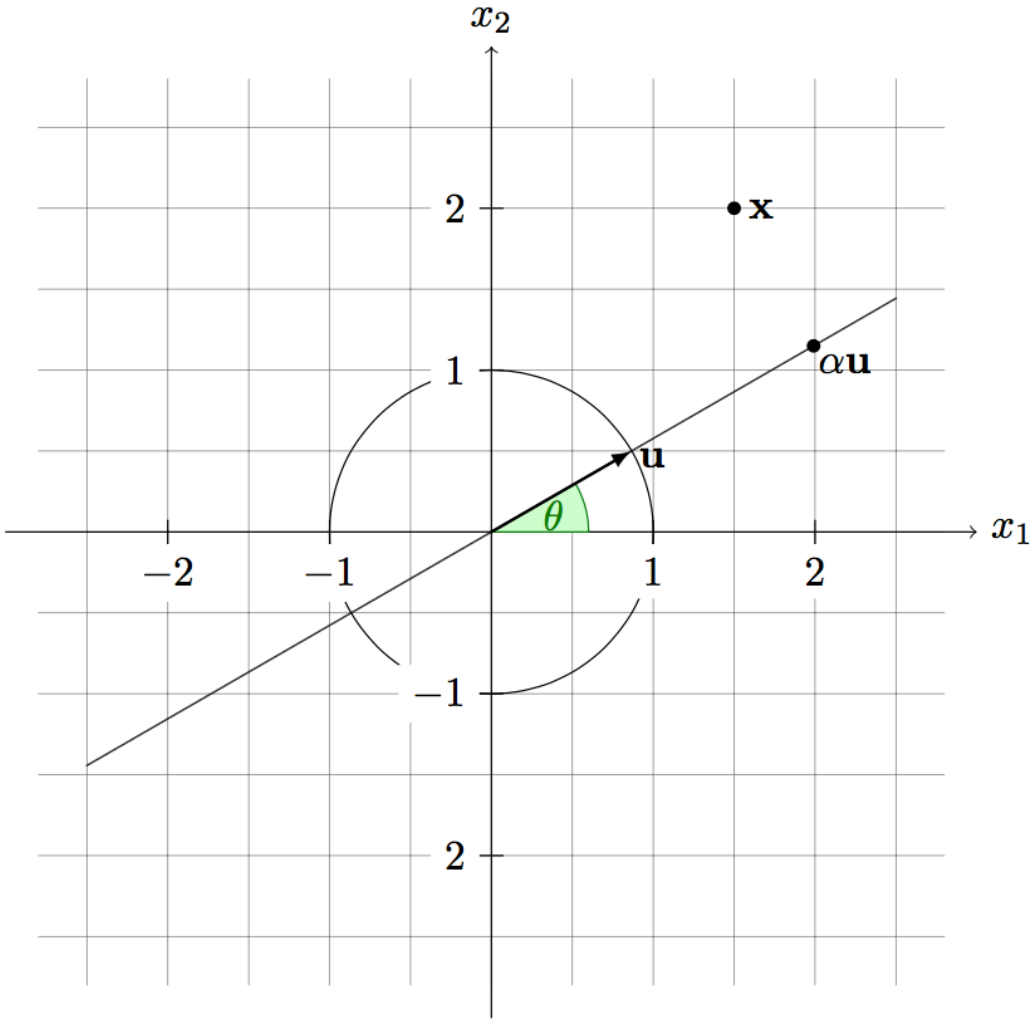
Exercise from Goodfellow et al Book Web Page

Slides by Anthony Maida

The expression $\alpha \boldsymbol{u}$ for $\alpha \in \mathbb{R}$ and unit vector $\boldsymbol{u} \in \mathbb{R}^n$ defines a line of points that may be obtained by varying the value of α. Derive an expression for the point $\boldsymbol{y}$ that lies on this line that is as close as possible to an arbitrary point $\boldsymbol{x} \in \mathbb{R}^n$. This operation of replacing a point by its nearest member within some set is called *projection*.

Possible application in data science: If you do an eigendecomposition on a dataset sampled from a bivariate distribution and want to project all data points onto the principle axis, this is a way to do it.

Diagram in 2D to Visualize the Problem



$\mathbf{x} = (1.5, 2.0)$
$\mathbf{u} = (.866, .5) = (1, \pi/6)$
$\theta = 30^\circ$ ($\pi/6$ rads)
$\alpha = 2.299$
$\alpha\mathbf{u} = (1.99, 1.15)$

Steps in Solution

Minimize

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2.$$

$$\|\boldsymbol{x} - \alpha\boldsymbol{u}\|^2.$$

$$\mathbf{x} - \alpha\mathbf{u} = \begin{bmatrix} x_1 - \alpha u_1 \\ x_2 - \alpha u_2 \\ \vdots \\ x_n - \alpha u_n \end{bmatrix}$$

$$\begin{aligned} &\|\boldsymbol{x} - \alpha\boldsymbol{u}\|^2 \\ =&(\boldsymbol{x} - \alpha\boldsymbol{u})^\top(\boldsymbol{x} - \alpha\boldsymbol{u}) \\ =&\boldsymbol{x}^\top\boldsymbol{x} - 2\alpha\boldsymbol{x}^\top\boldsymbol{u} + \alpha^2\boldsymbol{u}^\top\boldsymbol{u} \\ =&\boldsymbol{x}^\top\boldsymbol{x} - 2\alpha\boldsymbol{x}^\top\boldsymbol{u} + \alpha^2. \end{aligned}$$

This step explained on next slide.

Euclidean Distance of Point or Vector from Origin

$$||\mathbf{v}||$$

$$\sqrt{\sum_{i=1}^{n}(v_i - 0)^2} = \sqrt{\sum_{i=1}^{n} v_i^2}$$

$$||\mathbf{x} - \alpha\mathbf{u}||^2 = (\mathbf{x} - \alpha\mathbf{u})^T(\mathbf{x} - \alpha\mathbf{u})$$

Move transpose inward.

$$(\mathbf{x} - \alpha\mathbf{u})^T = \mathbf{x}^T - \alpha\mathbf{u}^T$$

$$||\mathbf{x} - \alpha\mathbf{u}||^2 = (\mathbf{x}^T - \alpha\mathbf{u}^T)(\mathbf{x} - \alpha\mathbf{u})$$

1. Multiplication is associative for matrices.

2. Addition and multiplication and are distributive for matrices.

3. Multiplication is **not** commutative for matrices. For instance, if a and b where matrices then ab would not necessarily equal ba. There are exceptions. For instance, since $\mathbf{x}$ and $\mathbf{u}$ are vectors, $\mathbf{u}^T\mathbf{x} = \mathbf{x}^T\mathbf{u}$.

$$(\mathbf{x}^T - \alpha\mathbf{u}^T)(\mathbf{x} - \alpha\mathbf{u}) = \mathbf{x}^T\mathbf{x} - \alpha\mathbf{u}^T\mathbf{x} - \alpha\mathbf{u}^T\mathbf{x} + \alpha^2\mathbf{u}^T\mathbf{u}$$

Formula for Distance as a Function of Alpha

$$f(\alpha) = \mathbf{x}^T\mathbf{x} - 2\alpha\mathbf{u}^T\mathbf{x} + \alpha^2$$

Notice how transpose creates a row vector.

$$f(\alpha) = [x_1, x_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 2\alpha\,[u_1, u_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \alpha^2$$

$$f(\alpha) = x_1^2 + x_2^2 - 2\alpha(x_1u_1 + x_2u_2) + \alpha^2$$

Take Derivative and Set to Zero

$$f(\alpha) = x_1^2 + x_2^2 - 2\alpha(x_1u_1 + x_2u_2) + \alpha^2$$

From previous slide

Derivative wrt alpha

$$f'(\alpha) = -2(x_1u_1 + x_2u_2) + 2\alpha$$

$$f'(\alpha) = -2\mathbf{x}^T\mathbf{u} + 2\alpha$$

Set to zero

$$f'(\alpha) = -2\mathbf{x}^T\mathbf{u} + 2\alpha = 0$$

$$\alpha = \mathbf{x}^T\mathbf{u} = [x_1, x_2]\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Solve for alpha