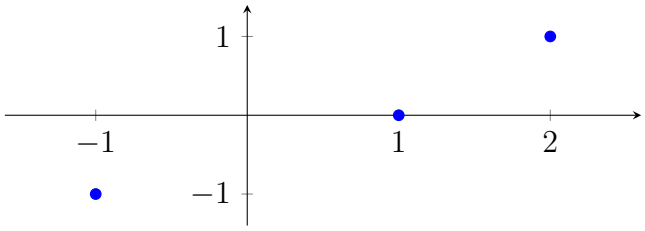
Least Squares Regression **Math 342**

Suppose we have three data points with x and y coordinates given by the following table:

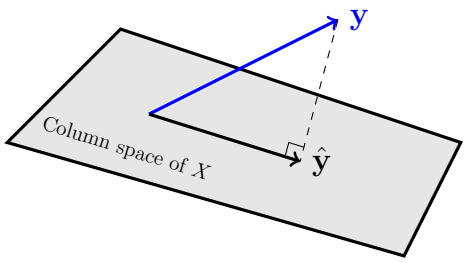
x	y
-1	-1
1	0
2	1



We want to find the linear function $\hat{y} = b_0 + b_1 x$ that is the best fit trendline for the scatterplot. If we **vectorize** this equation, that is treat $\mathbf{x}$ and $\mathbf{y}$ as vectors containing all of the x and y-values, then $\hat{\mathbf{y}} = X\mathbf{b}$ where

$$X = \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \\ 1 & x_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}.$$

The goal is to find the value of $\hat{\mathbf{y}}$ that is the closest to $\mathbf{y}$. Since $\hat{\mathbf{y}}$ is in the column space of X, we should try to find coefficients b_0 and b_1 such that $\hat{\mathbf{y}} - \mathbf{y}$ is orthogonal to the column space of X.



By the Fundamental Theorem of Linear Algebra, the orthogonal complement of the column space of X is the null space of X^T, so we must have

$$X^T(\hat{\mathbf{y}} - \mathbf{y}) = 0.$$

By substituting $X\mathbf{b}$ for $\hat{\mathbf{y}}$ and rearranging terms, we get the **normal equation** for linear regression:

$$X^T X\mathbf{b} = X^T\mathbf{y}.$$

1. Calculate $X^T X$ and $X^T\mathbf{y}$ for the example above.

2. Use linear algebra to solve the normal equation and find $\mathbf{b}$.

3. Use Matlab/Octave to find the best fit parabola $\hat{y} = b_0 + b_1 x + b_2 x^2$ for the four points $(-2, 3)$, $(-1, 0)$, $(1, -1)$, and $(3, 2)$. Start with the normal equations for the 4-by-3 Vandermonde matrix

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}.$$

In Matlab/Octave, you can enter the Vandermonde matrix using the following command:

```
X = fliplr(vander(x, n))
```

where `x` is a vector of x-values and `n` is the number of columns you want. The function `fliplr` reverses the order of the columns of a matrix (flips left-right) which is necessary because the `vander` function uses the opposite ordering.