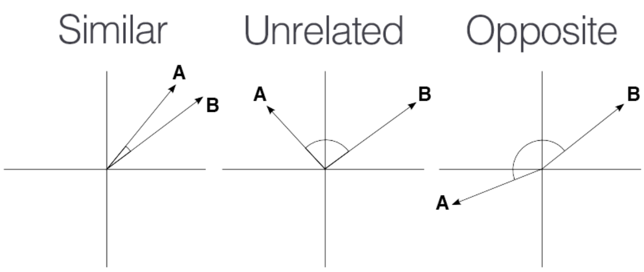
Linear Algebra

- Vector: $\mathbf{x} \in \mathbb{R}^{1\times D}$ (most of the times we will write $\mathbf{x} \in \mathbb{R}^{D}$ and mean the same thing)
  $\mathbf{x} = [x_1, \dots, x_D]$
- l_p norm:
  $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{D} |x_i|^p\right)^{1/p}$, $p \geq 1$
  $\|\mathbf{x}\|_0 = \sum_{i=1}^{D} \mathbb{I}\{x_i \neq 0\}$ (total number of non-zero elements in vector)
- Euclidean distance between two vectors $\mathbf{x_1} = [x_{11}, \dots, x_{1D}]$ and $\mathbf{x_2} = [x_{21}, \dots, x_{2D}]$:
  $\|\mathbf{x_1} - \mathbf{x_2}\|_2 = \sqrt{|x_{11} - x_{21}|^2 + \dots |x_{1D} - x_{2D}|^2} = \sqrt{\left(\sum_{i=1}^{D} |x_{1i} - x_{2i}|^2\right)}$
- Inner product between vectors $\mathbf{x_1} = [x_{11}, \dots, x_{1D}]$ and $\mathbf{x_2} = [x_{21}, \dots, x_{2D}]$:
  $< \mathbf{x_1}, \mathbf{x_2} >= (\mathbf{x_1}, \mathbf{x_2}) = x_{11} \cdot x_{21} + \dots x_{1D} \cdot x_{2D} = \|\mathbf{x_1}\|_2 \cdot \|\mathbf{x_2}\|_2 \cdot cos(\theta) \in \mathbb{R}$
  where θ is the angle between the vectors
- Cosine similarity (or angle θ) between vectors $\mathbf{x_1} = [x_{11}, \dots, x_{1D}]$ and $\mathbf{x_2} = [x_{21}, \dots, x_{2D}]$:
  $cos(\theta) = \frac{<\mathbf{x_1}, \mathbf{x_2}>}{\|\mathbf{x_1}\|_2 \|\mathbf{x_2}\|_2} \in [-1, 1]$
- Matrix: $\mathbf{X} \in \mathbb{R}^{D\times N}$, e.g., $\mathbf{X} = [\mathbf{x_1^T}, \dots, \mathbf{x_N^T}] = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{N1} \\ & \vdots & & \\ x_{1D} & x_{2D} & \dots & x_{ND} \end{bmatrix}$,
  where $\mathbf{x_i} = [x_{i1}, \dots, x_{iD}] \in \mathbb{R}^D$



- Vector-matrix multiplication ($\mathbf{X} \in \mathbb{R}^{D\times N}$, $\mathbf{w} \in \mathbb{R}^{1\times D}$):
  $$\mathbf{wX} = \underbrace{\begin{bmatrix} w_1 & \dots & w_D \end{bmatrix}}_{1\times D} \times \underbrace{\begin{bmatrix} x_{11} & x_{21} & \dots & x_{N1} \\ & \vdots & & \\ x_{1D} & x_{2D} & \dots & x_{ND} \end{bmatrix}}_{D\times N} = \underbrace{\begin{bmatrix} \mathbf{w}^T\mathbf{x_1} \\ \vdots \\ \mathbf{w}^T\mathbf{x_N} \end{bmatrix}}_{1\times N} \in \mathbb{R}^{1\times N}$$
- Gradient (or differential operator):
  $f : \mathbb{R}^D \to \mathbb{R}$, $\nabla f(\mathbf{x}) = \left[\frac{\theta f(\mathbf{x})}{\theta x_1}, \dots, \frac{\theta f(\mathbf{x})}{\theta x_D}\right] \in \mathbb{R}^D$
  e.g., $f(\mathbf{x}) = f(x_1, x_2) = x_1^2 + x_2^2$, $\nabla f(\mathbf{x}) = [2x_1, 2x_2] \in \mathbb{R}^D$

- Hessian matrix:
  $$f : \mathbb{R}^D \to \mathbb{R},\, \mathbf{H} = \nabla\left((\nabla f(\mathbf{x}))^T\right) = \begin{bmatrix} \frac{\theta^2 f(\mathbf{x})}{\theta x_1^2} & \cdots & \frac{\theta^2 f(\mathbf{x})}{\theta x_1 \theta x_D} \\ & \vdots & \\ \frac{\theta^2 f(\mathbf{x})}{\theta x_D \theta x_1} & \cdots & \frac{\theta^2 f(\mathbf{x})}{\theta x_D^2} \end{bmatrix} \in \mathbb{R}^{D\times D}$$
  e.g., $f(\mathbf{x}) = f(x_1, x_2) = x_1^2 + x_2^2 \in \mathbb{R}$, $\mathbf{H} = \nabla\left([2x_1, 2x_2]^T\right) = \nabla\left(\begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}\right) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$

- Basic matrix/vector derivatives:
  $\boldsymbol{\alpha}, \mathbf{x} \in \mathbb{R}^D$, $\mathbf{A} \in \mathbb{R}^{D\times D}$
  $\frac{\theta(\boldsymbol{\alpha}^T\mathbf{x})}{\theta\mathbf{x}} = \frac{\theta(\mathbf{x}^T\boldsymbol{\alpha})}{\theta\mathbf{x}} = \boldsymbol{\alpha}$
  $\frac{\theta^2(\mathbf{x}^T\mathbf{A}\mathbf{x})}{\theta\mathbf{x}^2} = \mathbf{A}$