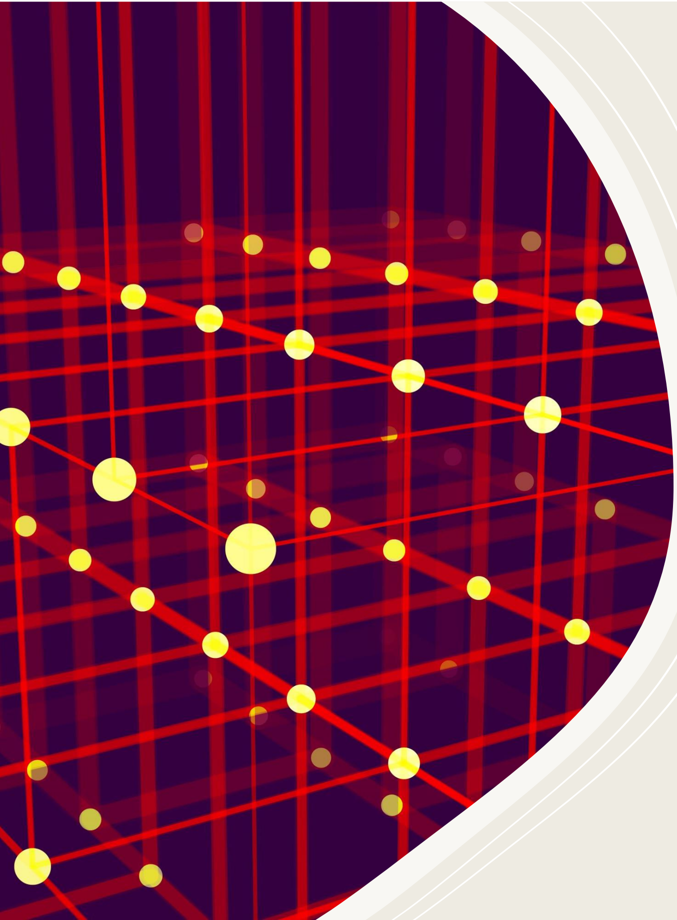


# CPSC 340 – Tutorial 4

Slides courtesy of Nam Hee Kim

University of British Columbia

# Agenda

- Optimization: Minimum and Minimizer
- Matrix/Vector/Norm Notation (Assign-related)
- Minimizing Quadratic Functions as Linear Systems (Assign-related)

# Optimization

Optimization, can simply be the art of making decisions that would give the best results.

- In optimization, you can either be minimizing  a function

$$min_x(x-a)^2+b$$

- to get a minimum value

- Or maximizing the function to get a maximum value

$$max_x(x-a)^2+b$$

- For every function, there is a minimum and its minimizer or a maximum and its maximizer

# Minimum and Minimizer

Given a function

$$f(x) = (sin(x - 0.5) + cos(x)^2) * 2$$

Minimize the function with respect to $x$ and return the **minimum** value and the **minimizer**

We can try to visualize this function given some input x values.

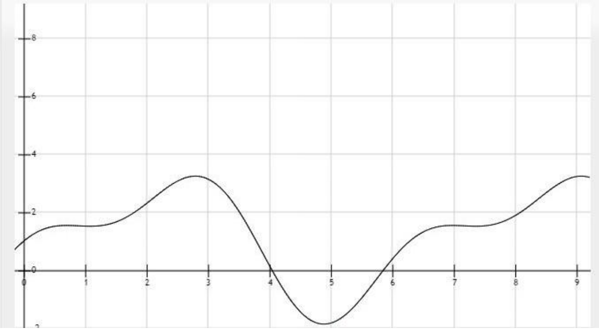
# Minimum and Minimizer

Given a function

$$f(x) = (sin(x - 0.5) + cos(x)^2) * 2$$

Minimize the function with respect to $x$ and return the **minimum** value and the **minimizer**

We can try to visualize this function given some input x values.



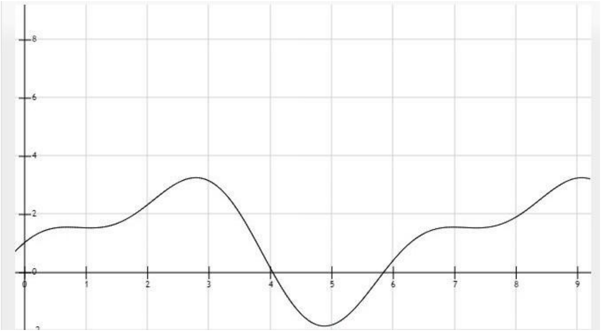
# Minimum and Minimizer

Given a function

$$f(x) = (sin(x - 0.5) + cos(x)^2) * 2$$

Minimize the function with respect to $x$ and return the **minimum** value and the **minimizer**

We can try to visualize this function given some input x values.




*What would the minimum and the minimizer ?*

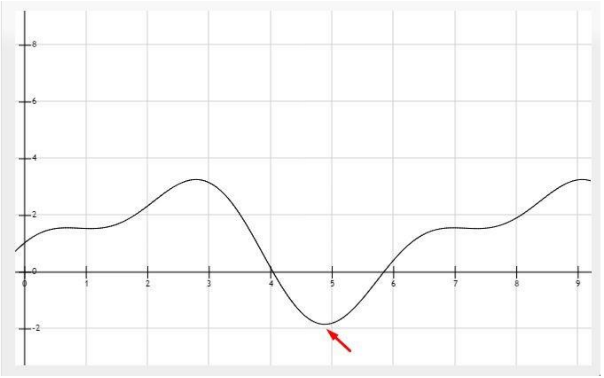
# Minimum and Minimizer

Given a function

$$f(x) = (sin(x - 0.5) + cos(x)^2) * 2$$

Minimize the function with respect to $x$ and return the **minimum** value and the **minimizer**

We can try to visualize this function given some input x values.



# What the code looks like

```python
def function(x):
    out = (np.sin(x-0.5) + np.cos(x)**2) * 2
    return out


x = np.array([0,1,2,3,4,5,6,7,8])

print('function f(x)', function(x))
```

```
function f(x) [ 1.04114892  1.54270424  2.34134635  3.15711457  0.15293351 -1.79413176
  0.43277331  1.56697719  1.91834047]
```

```python
'minimum', min(function(x))
```

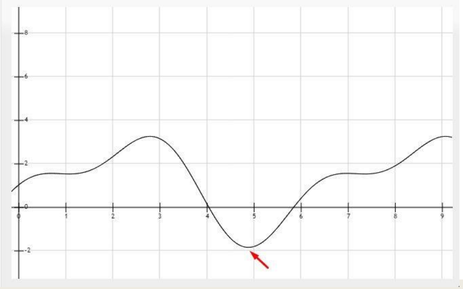
```
('minimum', -1.7941317644066466)
```

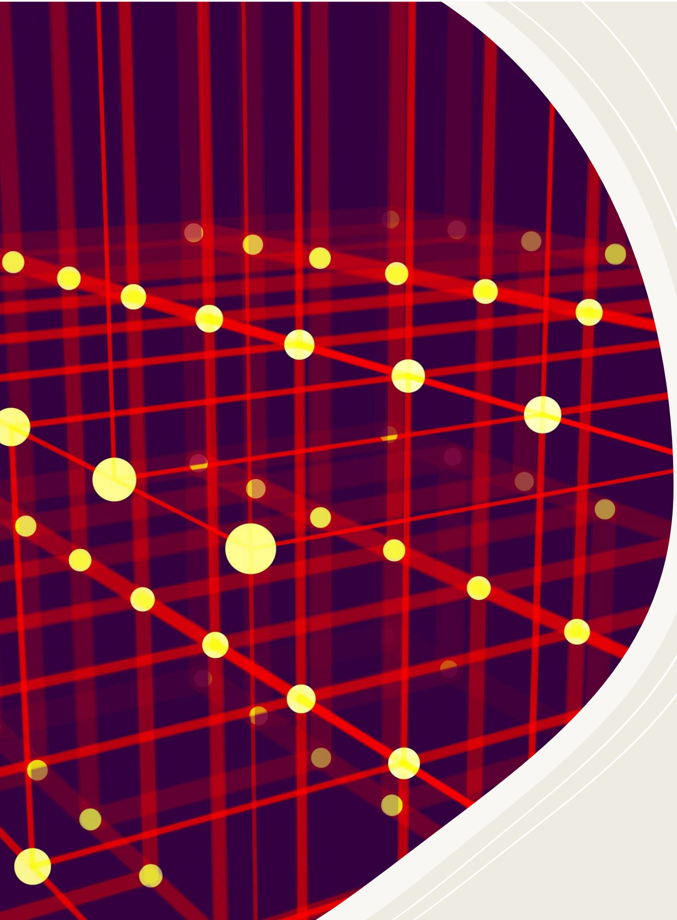
```python
'minimizer', x[np.argmin(function(x))]
```

```
('minimizer', 5)
```

```python
x = 5
function(x)
```

```
-1.7941317644066466
```





# Assignment Questions

# Matrix/Vector/Norm Notation

Express the following functions in terms of vectors, matrices, and norms (there should be no summations or maximums),

$$f(w) = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{\lambda}{2}\sum_{j=1}^{d} w_j^2$$

Recall, that all vectors are column-vectors,

$w_j$ is the scalar parameter $j$.
$y_i$ is the label of example $i$.
$x_i$ is the column-vector of features for example $i$.
$x_j^i$ is feature $j$ in example $i$.

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad x_i = \begin{bmatrix} x_1^i \\ x_2^i \\ \vdots \\ x_d^i \end{bmatrix}$$

# Matrix/Vector/Norm Notation

Let's first focus on the regularization term,

$$f(w) = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{\lambda}{2}\sum_{j=1}^{d} w_j^2$$

Recall the definition of inner product and L2-norm of vectors,

$$\|v\| = \sum_{j=1}^{d} v_j^2 \quad u^T v = \sum_{j=1}^{d} u_j v_j$$

Hence, we can write the regularizer in various forms using,

$$\|w\|^2 = \sum_{j=1}^{d} w_j^2 = \sum_{j=1}^{d} w_j w_j = w^T w$$

# Matrix/Vector/Norm Notation

Let's next focus on the least squares term,

$$f(w) = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{\lambda}{2}\sum_{j=1}^{d} w_j^2$$

Let's define the residual vector $r$ with elements

$$r_i = w^T x_i - y_i$$

We can write the least squares term as squared L2-norm of residual,

$j$

$$\sum_{i=1}^{n}(w^T x_i - y_i)^2 = \sum_{i=1}^{n} r_i^2 = r^T r = \|r\|^2$$

# Matrix/Vector/Norm Notation

Let's next focus on the least squares term,

$$f(w) = \frac{1}{2}\|r\|^2 + \frac{\lambda}{2}\|w\|^2, \quad r_i = w^T x_i - y_i$$

$X$ denotes the matrix containing the $x_i$ (transposed) in the rows: $X = \begin{bmatrix} — (x_1)^T — \\ — (x_2)^T — \\ \vdots \\ — (x_n)^T — \end{bmatrix}$

Using $w^T x_i = (x_i)^T w$ and the definitions of $r$, $y$, and $X$:

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} w^T x_1 - y_1 \\ w^T x_2 - y_2 \\ \vdots \\ w^T x_n - y_n \end{bmatrix} = \begin{bmatrix} (x_1)^T w \\ (x_2)^T w \\ \vdots \\ (x_n)^T w \end{bmatrix} - \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} — (x_1)^T — \\ — (x_2)^T — \\ \vdots \\ — (x_n)^T — \end{bmatrix}}_{X} w - y = Xw - y$$

$j$

Therefore $f(w) = \frac{1}{2}\|Xw - y\|^2 + \frac{\lambda}{2}\|w\|^2,$

# Minimizing Quadratic Function as a Linear System

A quadratic function is a function of the form

$$f(w) = \frac{1}{2} w^T A w + \ b^T w + y,$$

for a square matrix $A$, vector $b$, and scalar $y$.

Write the minimizer of the following function as a system of linear equations, using vector/matrix notation.

$$f(w) = \frac{1}{2} \sum_{i=1}^{n} (w^T x_i - y_i)^2 + \frac{\lambda}{2} \sum_{j=1}^{d} w_j^2$$

$j$ minimize convex functions, it is sufficient to find $w$ *s.t*

$f\,(w) = 0.$

# Minimizing Quadratic Function as a Linear System

Convert to vector/matrix form:

$$f(w) = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{\lambda}{2}\sum_{j=1}^{d} w_j^2 = \frac{1}{2}(Xw - y)^T(Xw - y) + \frac{\lambda}{2} w^T w$$

$$\rightarrow f(w) = \frac{1}{2} w^T X^T X w - w^T X^T y + \frac{1}{2} y^T y + \frac{\lambda}{2} w^T w$$

$j$

# Minimizing Quadratic Function as a Linear System

Convert to vector/matrix form:

$$f(w) = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{\lambda}{2}\sum_{j=1}^{d} w_j^2 = \frac{1}{2}(Xw - y)^T(Xw - y) + \frac{\lambda}{2} w^T w$$

$$\to f(w) = \frac{1}{2} w^T X^T X w - w^T X^T y + \frac{1}{2} y^T y + \frac{\lambda}{2} w^T w$$

Find $w$ such that $f'(w) = 0$:

$$f'_j(w) = X^T X w - X^T y + \lambda w = 0 \to (X^T X + \lambda I)w = X^T y$$

Note $f(w)$ is a column vector with dimension $d \times 1$.