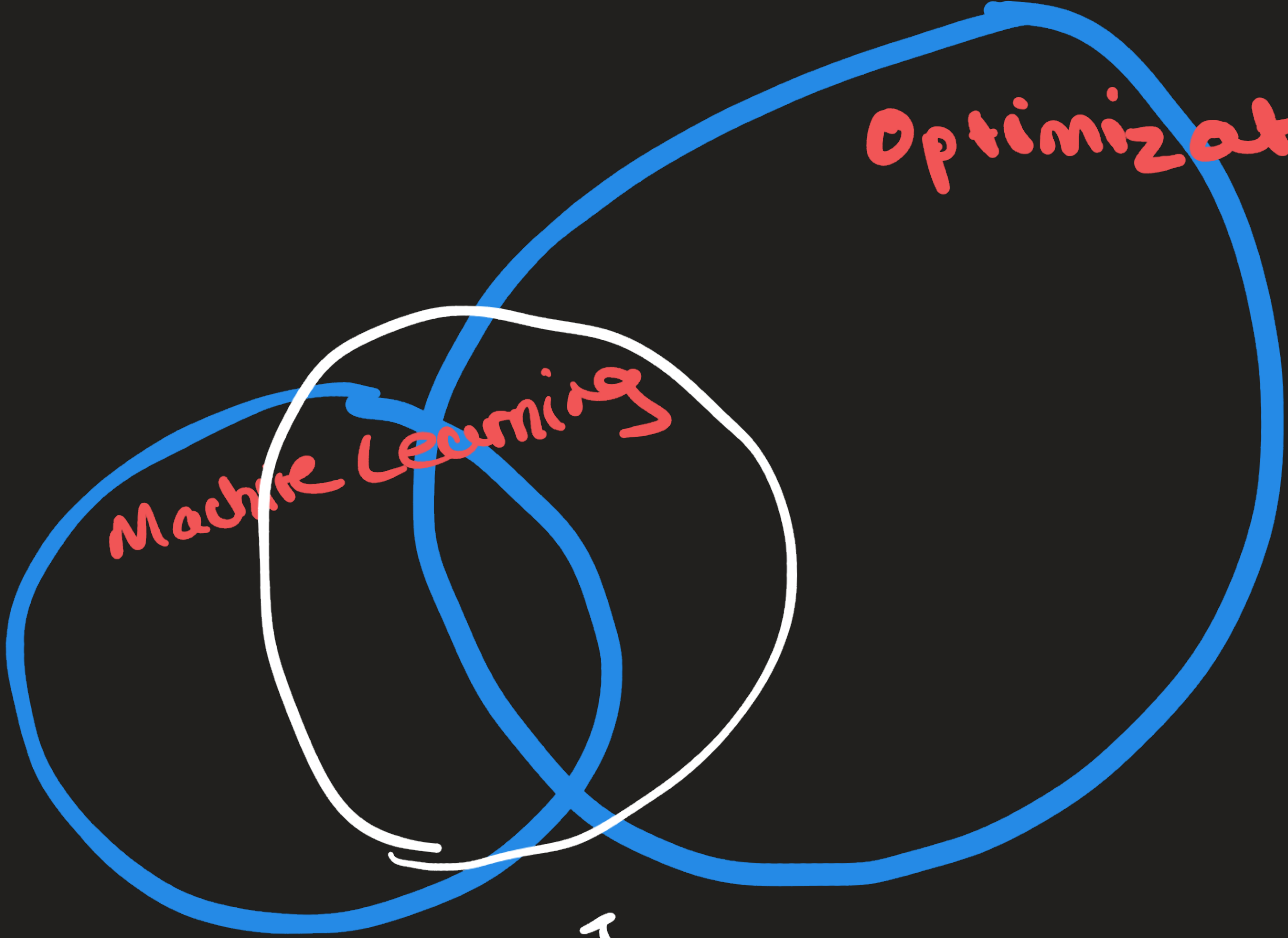

CSCI

4140/6962

What is Optimization?

- Finding extrema or extremal values of some objective

$$\theta^* \in \underset{\theta \in \Theta}{\operatorname{argmin}}\ J(\theta) \quad \text{or} \quad \theta^* \in \underset{\theta \in \Theta}{\operatorname{argmax}}\ J(\theta)$$

where

$\operatorname{argmin} J(\theta)$ denotes the set of parameters that achieve the <u>minimal</u> values

$\operatorname{argmax} J(\theta)$ denotes " <u>maximal</u> values

Examples

Least-squares regression:

$$Ax \approx b \quad \text{where } A \in \mathbb{R}^{n \times d},\ b \in \mathbb{R}^n$$

find x

assume salary $= x$ age

$$\begin{bmatrix} \text{salary}_1 \\ \vdots \\ \text{salary}_n \end{bmatrix} = x \begin{bmatrix} \text{age}_1 \\ \vdots \\ \text{age}_n \end{bmatrix}$$

Choose my objective that measures the goodness-of-fit:

$$J(x) = \sum_{i=1}^{n} (\text{salary}_i - x\, \text{age}_i)^2$$

Now we choose our model as

$$x^* \in \underset{x \in \mathbb{R}}{\operatorname{argmin}} \; J(x) = \sum_{i=1}^{n} (\text{salary}_i - x\, \text{age}_i)^2$$

Recall the Euclidean norm $\|z\|_2$ of a vector $z \in \mathbb{R}^n$ is given by

$$\|z\|_2^2 = \sum_{i=1}^{n} z_i^2$$

Write the residual, or error, vector

$$\overline{\text{salary}} - x\, \overline{\text{age}}$$

then

$$J(x) = \|\overline{\text{salary}} - x\, \overline{\text{age}}\|_2^2$$

More generally, if we predict salary using d predictors $z_1, \dots, z_d$, e.g. $z_1 = \text{age}, \dots, z_d = \text{latitude}$ then we can formulate the prediction problem as

$$J(x) = \left\| \overline{\text{salary}} - \begin{bmatrix} \bar{z}_1 & | & \cdots & | & \bar{z}_d \end{bmatrix} x \right\|_2^2$$

where $x \in \mathbb{R}^d$

Example

Nonnegative Least-Squares (e.g. chemometrics or an alternative to PCA)

$$x^* \in \underset{x \geq 0}{\operatorname{argmin}} \; \|Ax - b\|_2^2$$

Solving optimisation problems

$$x^* \in \operatorname{argmin} J(x)$$

minimizers satisfy $\nabla J(x) = 0$

where

$$[\nabla J(x)]_i = \frac{\partial J}{\partial x_i}(x)$$

Example of using optimality conditions to solve an optimization problem

Consider least-square

$$J(x) = \| Ax - b \|_2^2$$

$$\nabla J(x) = 2A^T(Ax-b) = 0$$

or equivalently

$$A^T A x = A^T b$$

so the solution to the LS-problem is

$$x = (A^T A)^{-1} A^T b \qquad \text{if } A^T A \text{ is invertible}$$

If A^TA is not invertible, then the minimum-norm solution is

$$x^+ = (A^TA)^+ A^Tb,$$

where $(A^TA)^+$ is the pseudoinverse of A^TA

In general $\nabla J(x) = 0$ is a nonlinear set of equations, so we don't have a nice closed-form solution. Instead, we use gradient descent or similar numerical optimization procedure.

Gradient Descent

x_0 = some initial point

for $t = 0, \dots, T-1$

$$x_{t+1} = x_t - \mu \nabla J(x_t)$$ where μ is the step-size

Under some assumptions on J, μ, T this process converges to a minimizer of J.

Example of the poor behavior of gradient descent

$$x = \arg\min \frac{1}{2} \| Dx \|_2^2 \qquad D = \begin{bmatrix} 2 & 0 \\ 0 & 50 \end{bmatrix}$$

(see code linked from website)