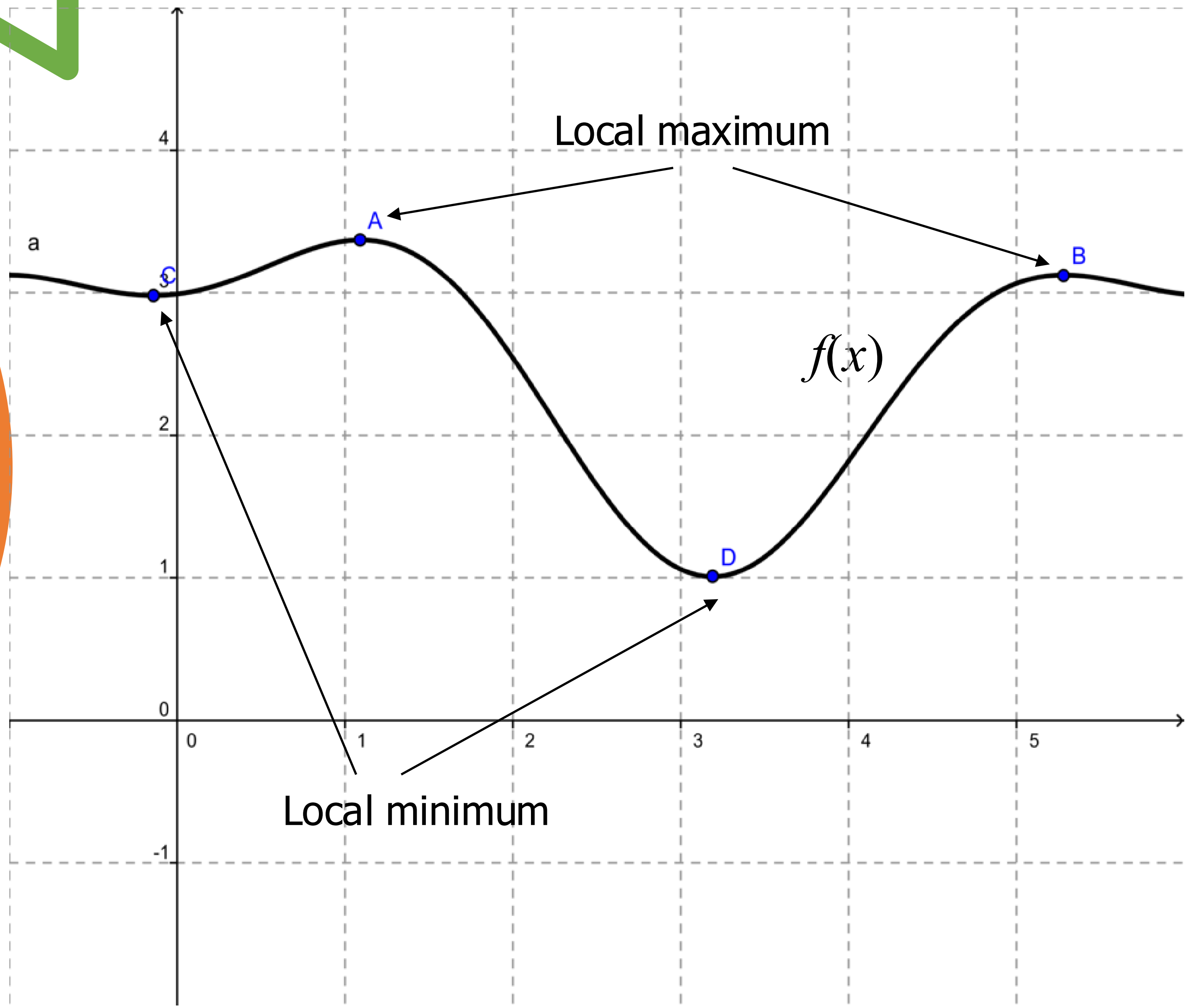
ML-5

Machine Learning

- Supervised Learning
  - Lagrange based Classification
    - Support Vector Machines

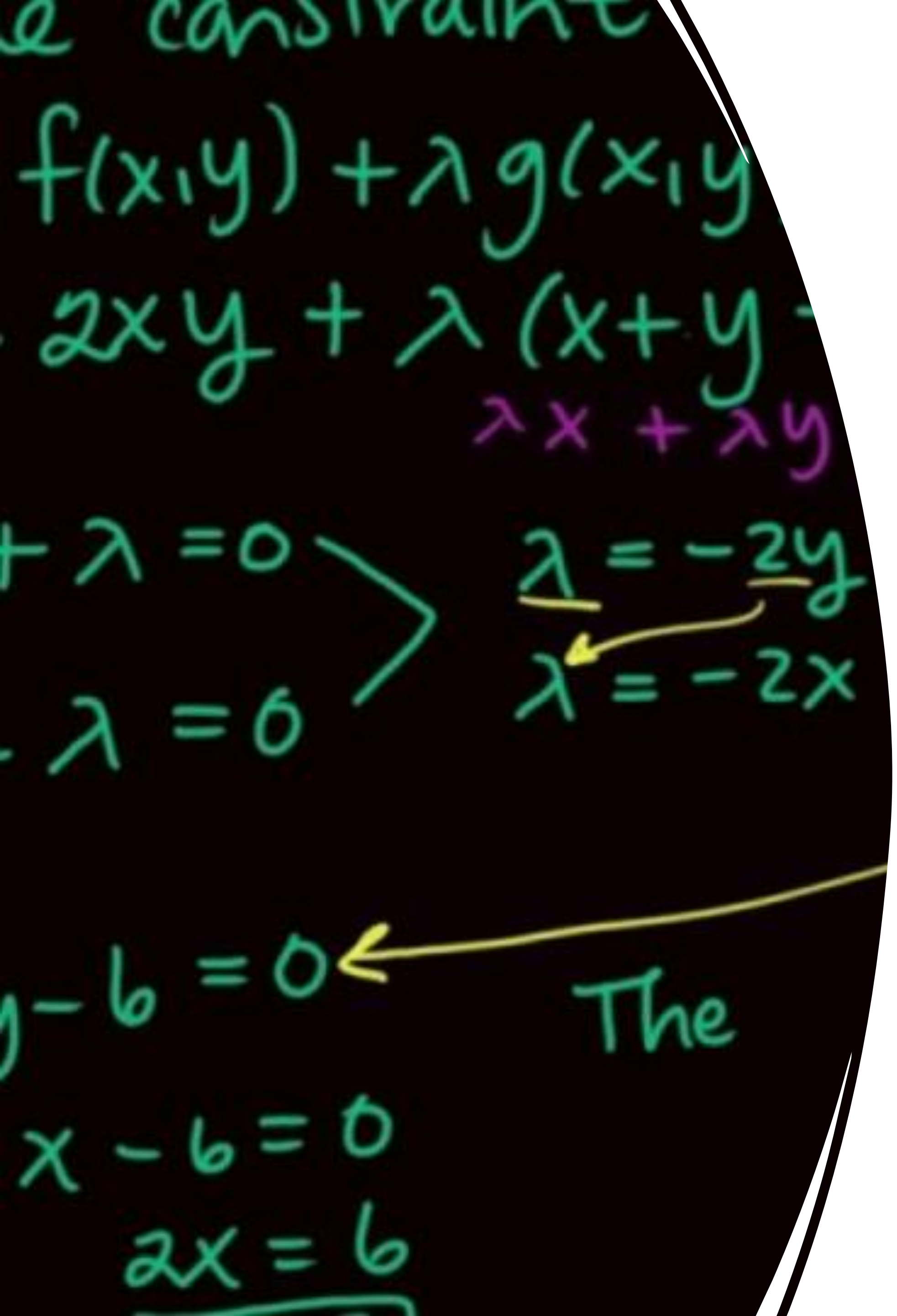
Prof. Umut ORHAN

Optimization



Joseph-Louis Lagrange

«Mécanique Analytique» (1788)



$$L(x,\alpha) = f(x) + \alpha\, g(x)$$ 1

$$\frac{\partial L(x,\alpha)}{\partial x} = 0$$

$$\frac{\partial L(x,\alpha)}{\partial \alpha} = 0$$ 2

Lagrange Optimization

Example

Find the optimum value of the $f(x, y) = 25 - x^2 - y^2$ satisfying the constraint $4 - x - y = 0$

$$L(x,y,\alpha) = 25 - x^2 - y^2 + \alpha(4 - x - y)$$

$$\frac{\partial L(x,y,\alpha)}{\partial x} = -2x - \alpha = 0$$

$$\frac{\partial L(x,y,\alpha)}{\partial y} = -2y - \alpha = 0$$

$$\frac{\partial L(x,y,\alpha)}{\partial \alpha} = 4 - x - y = 0$$

$$x = -0.5\alpha \qquad \alpha = -4$$

$$y = -0.5\alpha \qquad x = 2$$

$$x + y = 4 \qquad y = 2$$

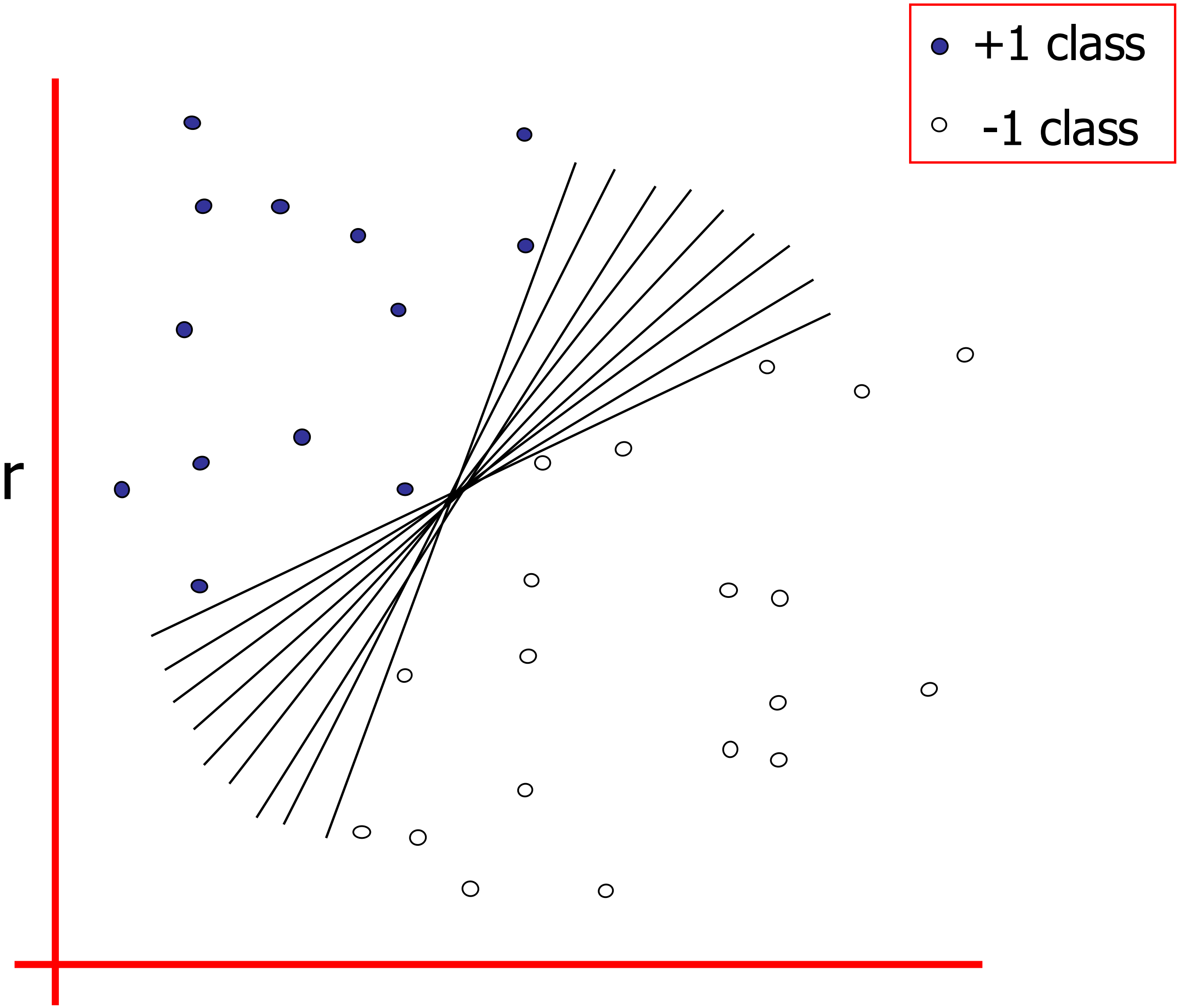
$$f(x = 2, y = 2) = 25 - 2^2 - 2^2 = 17$$

Vladimir Vapnik

Support Vector Machines (1992)

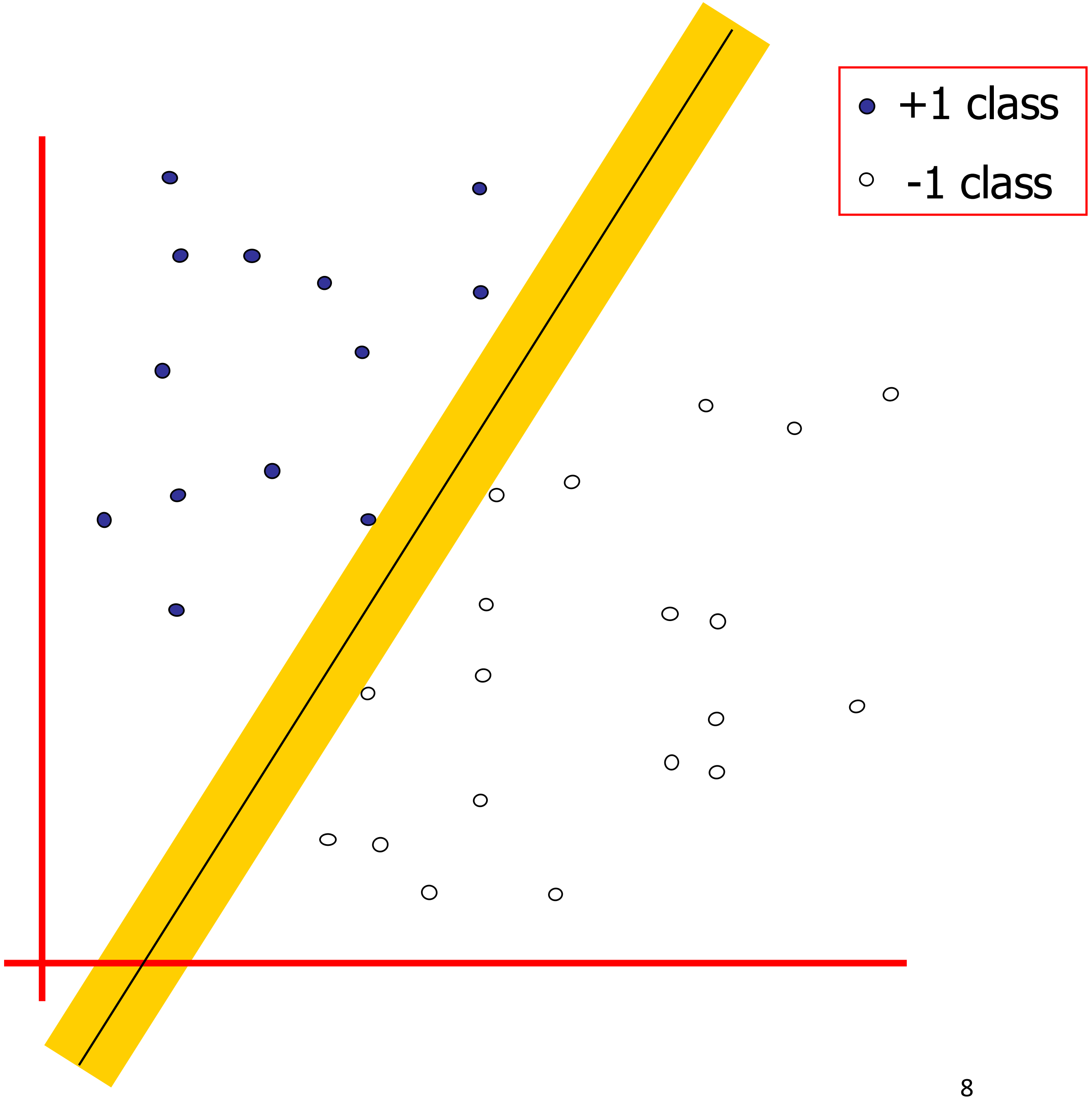
Support Vector Machine (SVM)

Which one is the best separator line for the classes?



Support Vector Machine (SVM)

The line that maximizes the margin between two classes is the best separator line.



Support Vector Machine (SVM)

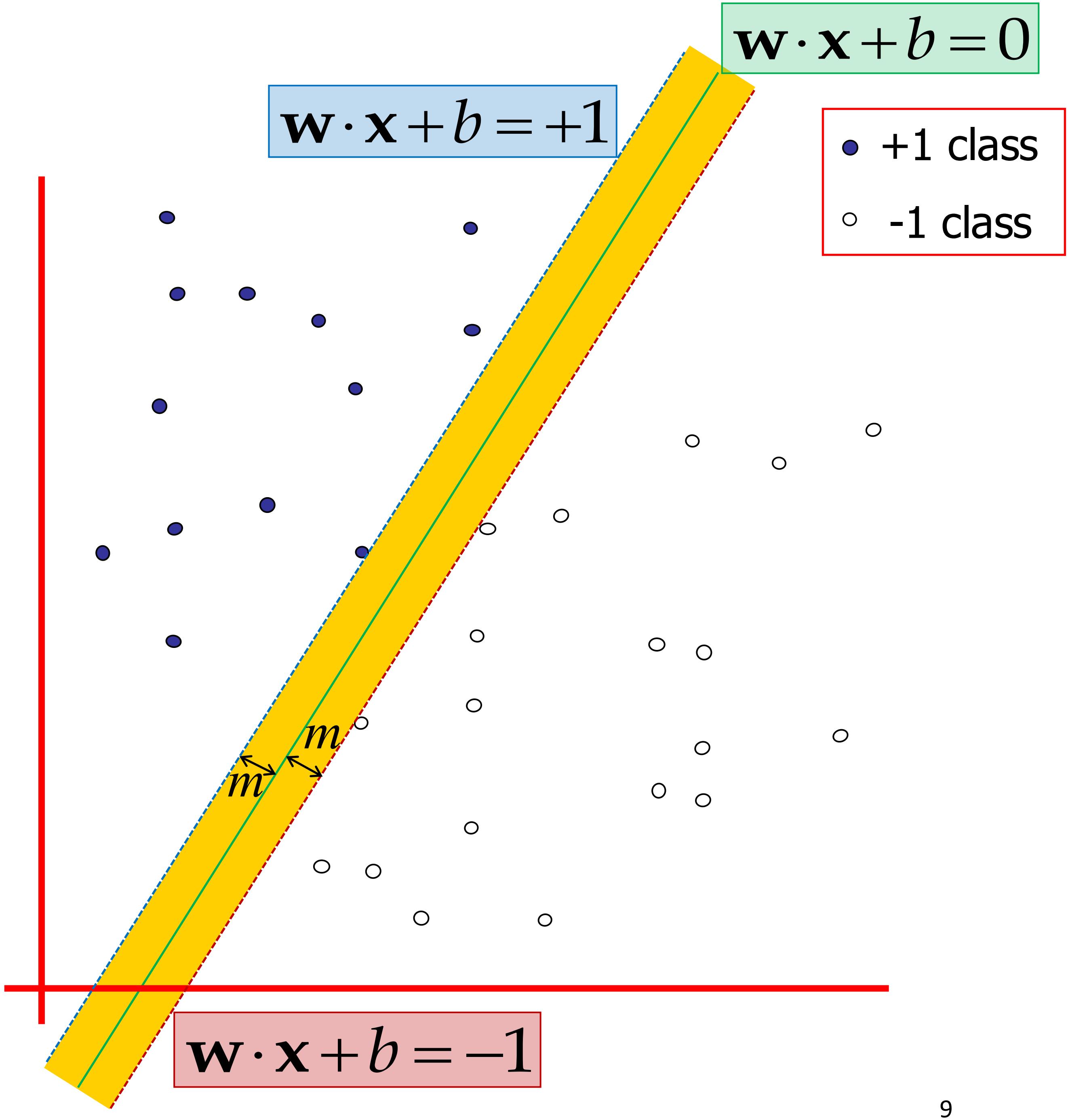
$$w \cdot x_i + b \geq +1 \text{ for } d_i = +1$$

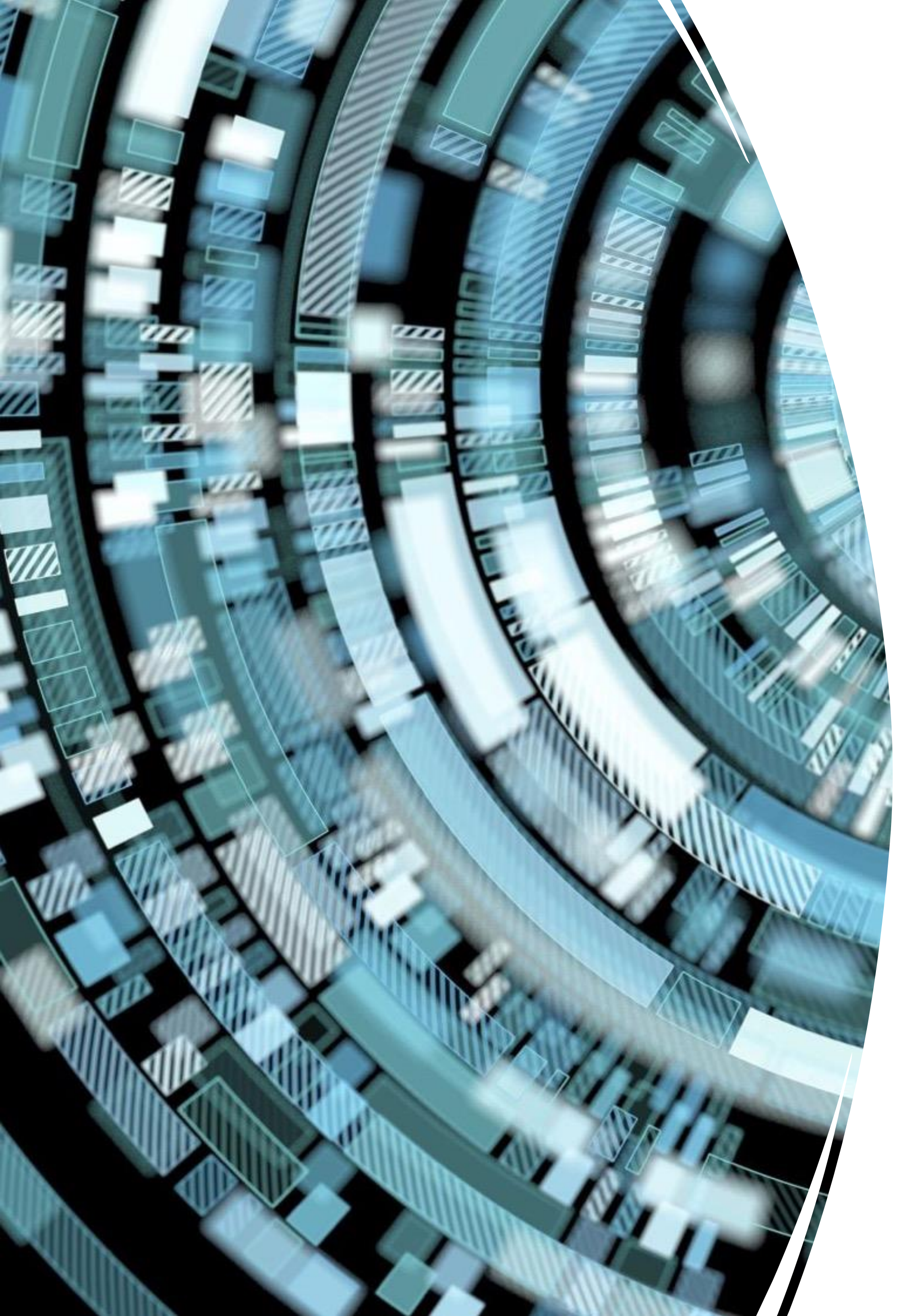
$$w \cdot x_i + b \leq -1 \text{ for } d_i = -1$$

$$m = \frac{2}{\sqrt{w \cdot w}} \rightarrow f_{\min}(w) = \frac{w \cdot w}{2}$$

$$d_i(w \cdot x_i + b) \geq 1, \quad 1 \leq i \leq N$$

$$L(w, b, \alpha) = \frac{w.w}{2} - \sum_i \alpha_i \left(d_i(w \cdot x_i + b) - 1\right)$$





Support Vector Machine (SVM)

$$\frac{\partial L(w,b,\alpha)}{\partial w} = 0 \rightarrow \qquad w = \sum_i \alpha_i d_i x_i$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = 0 \rightarrow \qquad \sum_i \alpha_i d_i = 0$$

$$L(w,b,\alpha) = \frac{w.w}{2} - \sum_i \alpha_i \left(d_i \left(w \cdot x_i + b \right) - 1 \right)$$

$$L(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j d_i d_j x_i x_j$$

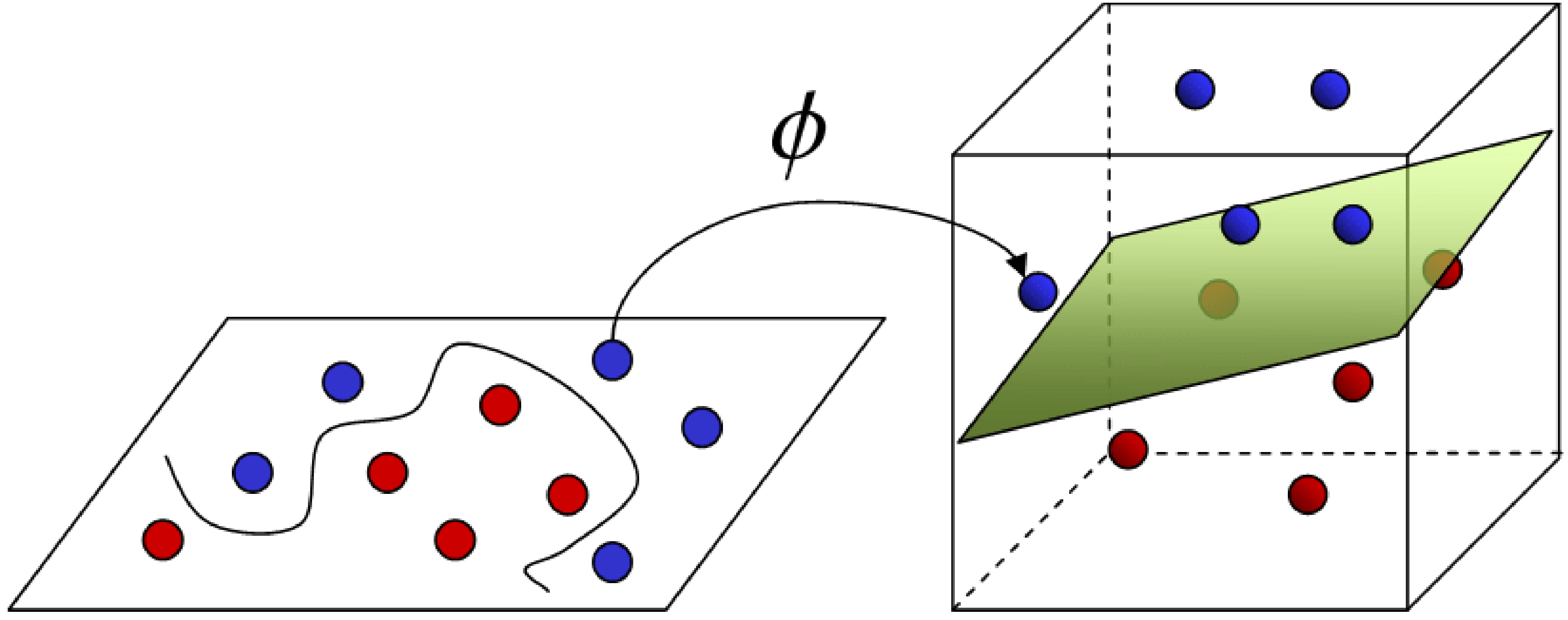
$$L(\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j d_i d_j x_i x_j$$

$$R \approx -D^T D X^T X \qquad and \qquad u \approx \alpha_i$$

$$\min_u \left(\frac{u^T R u}{2} + d^T u + c \right)$$

Quadratic term

Quadratic Programming



Mapping by Kernel

A kernel function can transform 2-D data into a 3-D space. This enables the classes to be linearly separated from each other.

$$K(x_i, x_j) = \begin{cases} x_i x_j & \textit{linear} \\ (1 + x_i x_j)^p & \textit{polynomial} \\ \exp\left(-\frac{1}{2}\left\|x_i - x_j\right\|^2\right) & \textit{RBF} \end{cases}$$

Mapping by Kernel

The choice of kernel function is based on the characteristics of the data and is adjusted accordingly to enhance classification performance.

Advantages

SVMs hold significant importance in ML for various reasons:

1. SVMs can achieve high accuracy rates even with non-linear datasets.

2. SVMs maximize the margin to prevent the model from overfitting.

3. SVMs can be used for many machine learning problems, including classification, regression.

4. SVMs can provide effective results not only with large datasets but also with small ones.

5. SVMs are grounded in statistical learning theory, making them well-defined and understood theoretically.

Machine Learning

5. week

Thanks for watching