Johns Hopkins Engineering

Applied Machine Learning for Mechanical Engineers

Popular Supervised Machine Learning Techniques, Part 1, A

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Extreme Learning Machine & Support Vector Machine

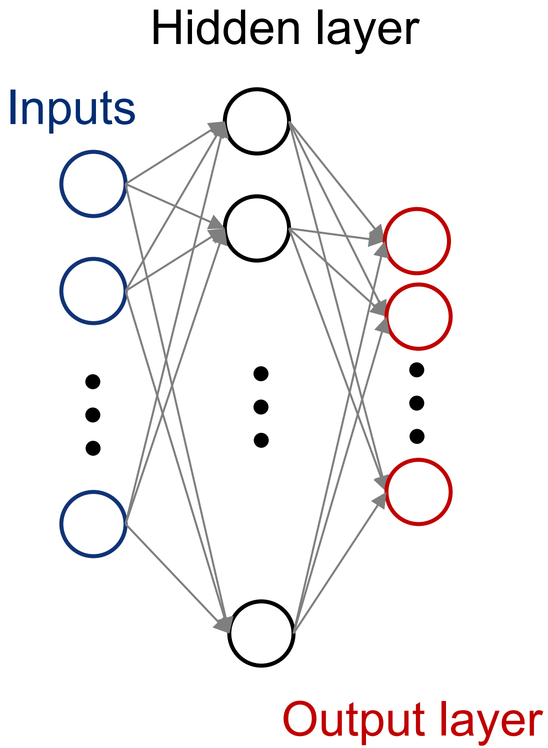
- By the end of this lecture you will be able to:

  - Describe Extreme Learning Machine (ELM)

  - Describe Support Vector Machine (SVM)

Extreme Learning Machine & Support Vector Machine

- Extreme Learning Machine* (ELM)
  - Three-Layer Neural Network
    - Input layer
    - Hidden layer
    - Output Layer
  - Similar to a shallow neural network
  - Both classification and regression problems

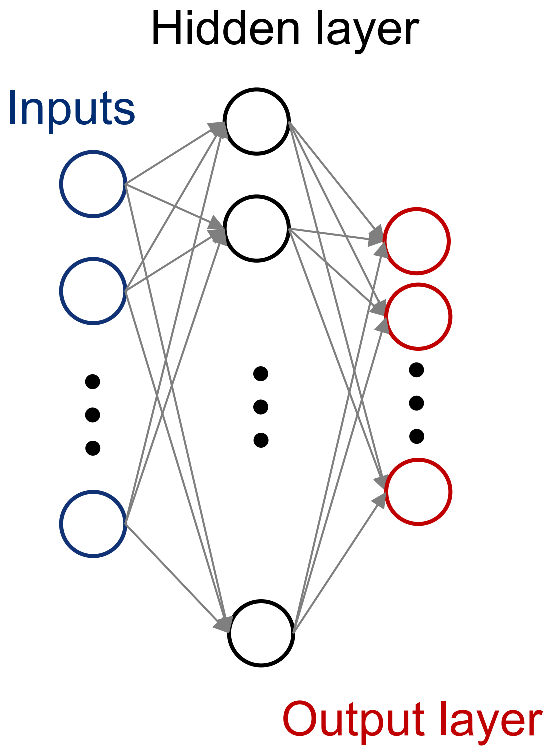


*Huang, G. B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: a survey. International journal of machine learning and cybernetics, 2(2), 107-122.

Extreme Learning Machine & Support Vector Machine

- Extreme Learning Machine (ELM)
  - Three-Layer Neural Network
    - $\boldsymbol{H} = a(\boldsymbol{X}, \boldsymbol{P})$
    - $\boldsymbol{O} = \boldsymbol{H}\,\boldsymbol{W}_{out}$

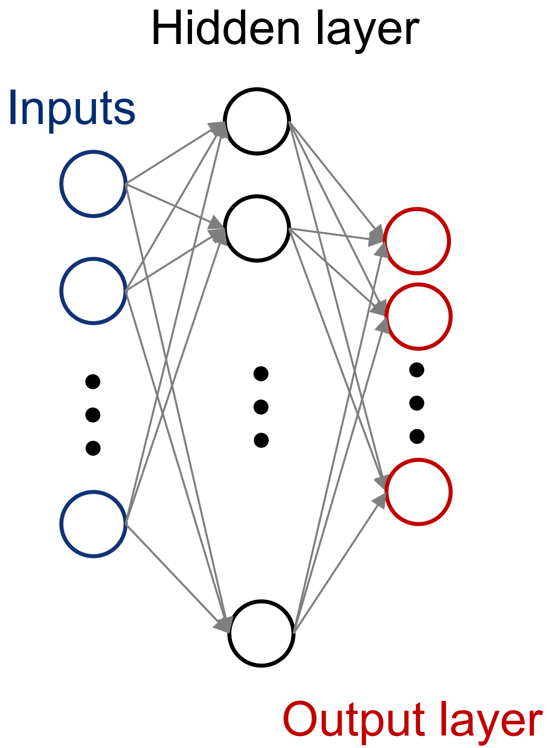
  where $\boldsymbol{X}$ is an I-dimensional row vector of inputs, $\boldsymbol{H}$ is an L-dimensional row vector of hidden neurons, $a(.)$ is an activation function (also called Kernel function), $\boldsymbol{P}$ is the set of parameters of the activation function, $\boldsymbol{W}_{out}$ is an L by M (M is the number of outputs) matrix of weight connections between hidden and output layers, and $\boldsymbol{O}$ is an M-dimensional row vector of outputs.



Extreme Learning Machine & Support Vector Machine

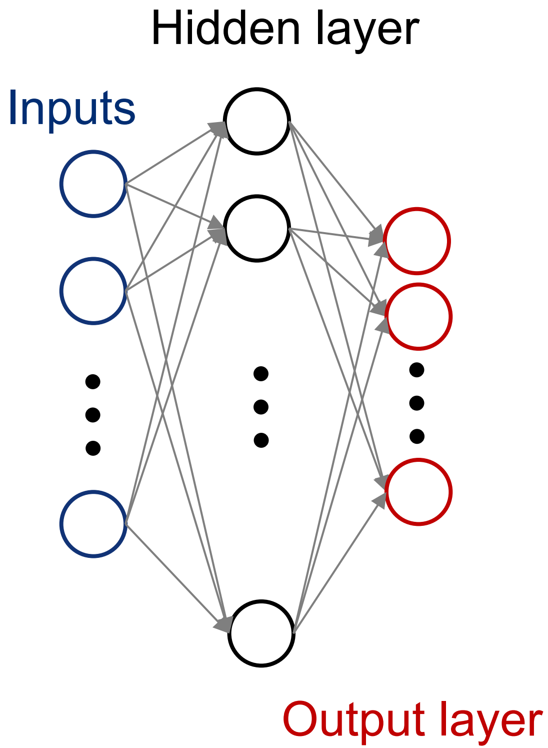
- Extreme Learning Machine (ELM)
  - Three-Layer Neural Network
    - $\boldsymbol{H} = a(\boldsymbol{X}, \boldsymbol{P})$
    - $\boldsymbol{O} = \boldsymbol{H}\,\boldsymbol{W}_{out}$
  - Example: A sigmoidal activation function
    - $\boldsymbol{P} = \{\boldsymbol{W}_{in}, \boldsymbol{B}\}$
    - $a(\boldsymbol{X}, \boldsymbol{P}) = \dfrac{1}{1+e^{-(\boldsymbol{X}\,\boldsymbol{W}_{in}+\boldsymbol{B})}}$

    Where $\boldsymbol{W}_{in}$ is an I by L matrix of weight connections between input and hidden layers, and $\boldsymbol{B}$ is an L-dimensional row vector of biases.



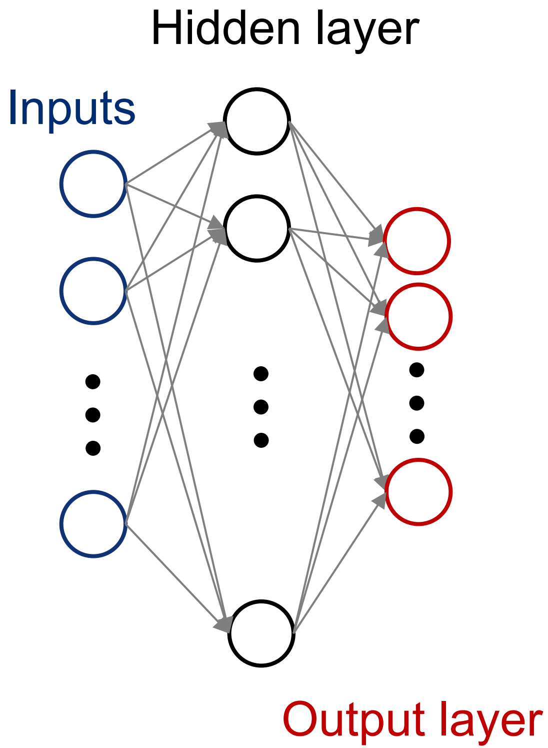
Extreme Learning Machine & Support Vector Machine

- Extreme Learning Machine (ELM)
  - Three-Layer Neural Network
    - $\boldsymbol{H} = a(\boldsymbol{X}, \boldsymbol{P})$ where $\boldsymbol{P} = \{\boldsymbol{W}_{in}, \boldsymbol{B}\}$
    - $\boldsymbol{O} = \boldsymbol{H}\,\boldsymbol{W}_{out}$
  - Parameters, $\boldsymbol{W}_{in}$ and $\boldsymbol{B}$, are generated randomly.



Extreme Learning Machine & Support Vector Machine

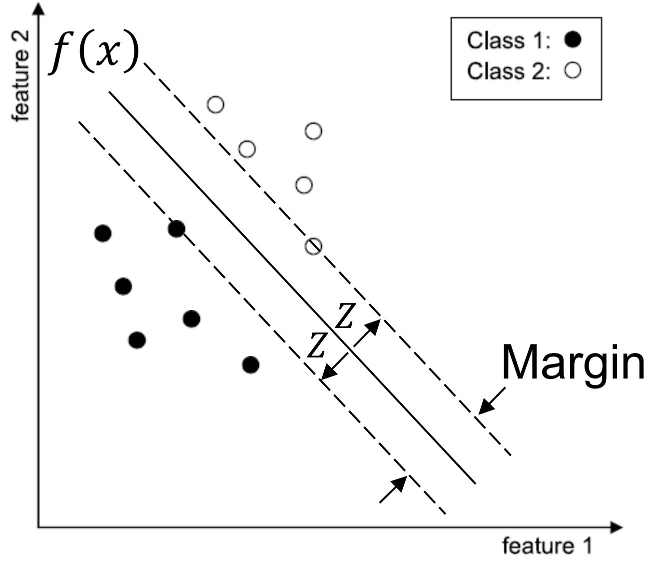
- Extreme Learning Machine (ELM)
  - Three-Layer Neural Network
    - $\boldsymbol{H} = a(\boldsymbol{X}, \boldsymbol{P})$ where $\boldsymbol{P} = \{\boldsymbol{W}_{in}, \boldsymbol{B}\}$
    - $\boldsymbol{O} = \boldsymbol{H}\,\boldsymbol{W}_{out}$
  - Training: Compute only $\boldsymbol{W}_{out}$ as follows
    - $\boldsymbol{W}_{out} = \widehat{\boldsymbol{H}}\boldsymbol{O}$
    - $\widehat{\boldsymbol{H}}$ is the Moore-Penrose inverse* of $\boldsymbol{H}$



*Rao CR, Mitra SK (1971) Generalized inverse of matrices and its applications. Wiley, New York

Extreme Learning Machine & Support Vector Machine

- Support Vector Machine (SVM)
  - Originally proposed for two-category classification problems
  - The goal is to find the best hyperplane, $f(\boldsymbol{x})$, that separates the data with the highest margin, Z, where $\boldsymbol{x}$ is the vector of features (e.g. inputs or outputs of a Kernel).
  - Both classification and regression problems

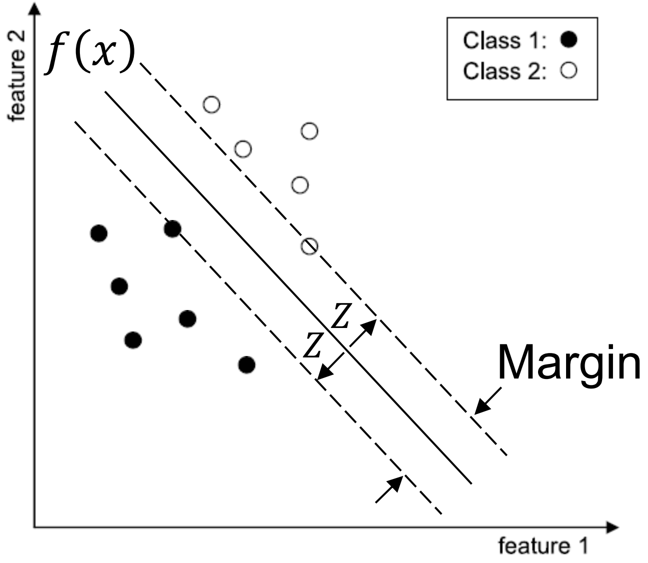


*Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Extreme Learning Machine & Support Vector Machine

- Support Vector Machine (SVM)
  - $f(\boldsymbol{x}) = \boldsymbol{W}^T\boldsymbol{x} + \boldsymbol{W_0}$
  - $-1 \le f(\boldsymbol{x}) \le +1$
  - $\begin{cases} f(\boldsymbol{x}) \ge +1, & \forall \boldsymbol{x} \in class\ 1 \\ f(\boldsymbol{x}) \le -1, & \forall \boldsymbol{x} \in class\ 2 \end{cases}$
  - $Z = \dfrac{|f(\boldsymbol{X})|}{\|\boldsymbol{W}\|}$

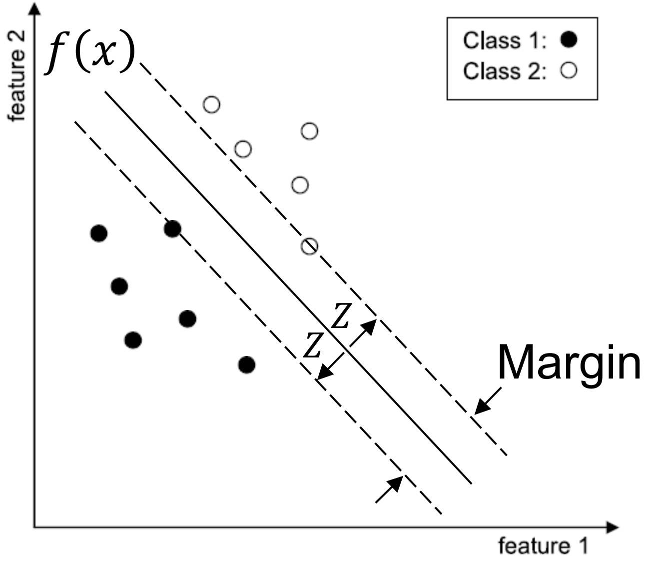
  where $\boldsymbol{W}$ is the hyperplane normal vector, $\boldsymbol{W_0}$ is the vector of hyperplane constants, and Z is the distance between the hyperplane and the nearest point $\boldsymbol{x} = \boldsymbol{X}$, and is so-called the "Support Vector"



*Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Extreme Learning Machine & Support Vector Machine

- Support Vector Machine (SVM)
  - Z is maximum when $|f(\boldsymbol{X})| = 1$
  - The margin, $2Z$, is maximum when $Z = \frac{1}{\|\boldsymbol{W}\|}$. The margin is then $2Z = \frac{2}{\|\boldsymbol{W}\|}$
  - To maximize the margin, $\boldsymbol{W}$ needs to be minimized

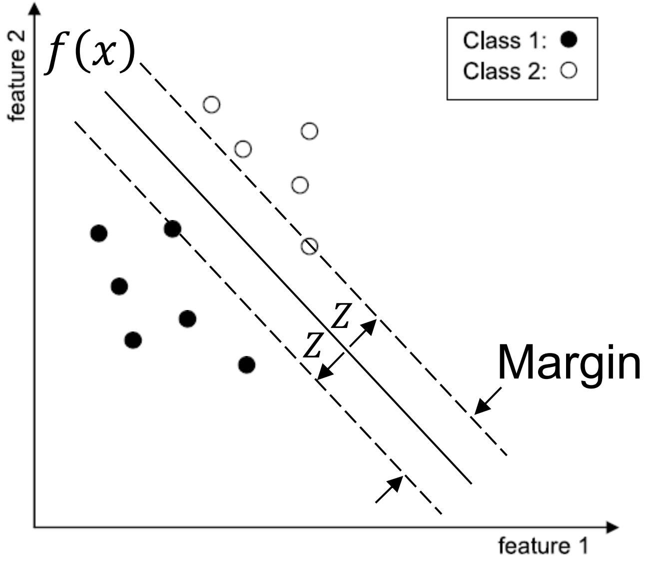


*Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Extreme Learning Machine & Support Vector Machine

- Support Vector Machine (SVM)
  - $\boldsymbol{W}$ can be defined as the linear combination of training vectors:
    - $\boldsymbol{W} = \sum_{i=0}^{N} \alpha_i Y_i \boldsymbol{X}_i$
    - Subject to $\sum_{i=0}^{N} \alpha_i Y_i = 0$

where N is the number of training samples, Y_i is the output corresponding to input vector $\boldsymbol{X}_i$, and α_i is the Lagrange multiplier corresponding to $\boldsymbol{X}_i$.



*Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Extreme Learning Machine & Support Vector Machine

- In this lecture, you learned about:
  - Extreme Learning Machine (ELM)
  - Support Vector Machine (SVM)
- In the next lecture, we will talk about probabilistic and enhanced probabilistic neural networks

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING