BITS F464: Machine Learning

24 SOM, VC-Dimension and Monte Carlo Simulation

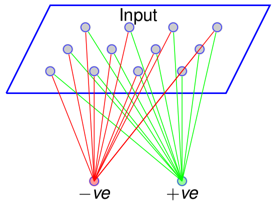
Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

March 19, 2021 ONLINE (Campus @ BITS-Pilani Jan-May 2021)

http://ktiwari.in/ml

Self-Organizing Map (SOM)

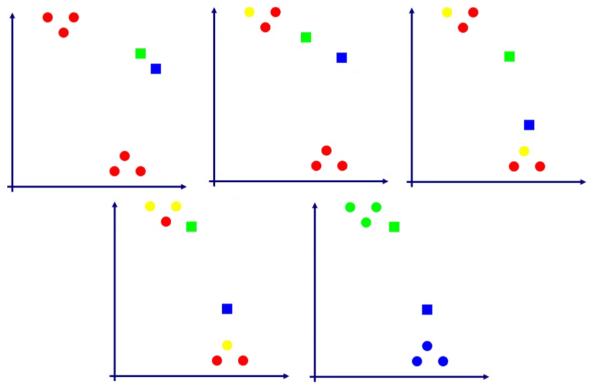
- SOM neural network models is based on unsupervised, competitive learning
- Provides a topology preserving mapping
- Neurons are represented using weight vector called the codebook $(w_1, w_2, .., w_n)$. Consider input as $x = (x_1, x_2, ..., x_n)$
- SOM can be used of clustering and feature detection (SOFM)
- We would see Kohonen model



Self-Organizing Map (SOM)

- Three Process
  1. **Competition:** discriminant function is computed
     $$argmin_j ||\vec{x} - \vec{w_j}||$$
  2. **Cooperation:** neuron in spacial neighborhood h_{ij} are excited
     $$h_{ij} = e^{-d_{ij}^2/(2\sigma^2)}$$
     d_{ij} is lateral distance between neuron i and j. σ decreases with every iteration $\sigma(n) = \sigma_0 e^{-n/\tau_1}$
  3. **Synaptic Adaptation:** weight adjustment (Hebbian learning)
     $$\triangle \vec{w_j} = \eta h_{ij}(\vec{x} - \vec{w_j})$$
     $$\vec{w_j}(n+1) = \vec{w_j}(n) + \eta(n) h_{ij}(\vec{x} - \vec{w_j})$$
     where $\eta(n) = \eta_0 e^{-n/\tau_2}$

Self-Organizing Map (SOM)



Computational Learning Theory

Is it possible to identify classes of learning problems that are inherently difficult or easy?

- **Sample complexity.** Number of training examples needed for a learner to converge with high probability to a successful hypothesis?
- **Probably approximately correct** (PAC) learning model

  Definition: Consider a concept class C defined over a set of instances X of length n and a learner L using hypothesis space H. C is PAC-learnable by L using H if for all $c \in C$, distributions D over X, we have ϵ and δ such that $0 < \epsilon, \delta < 1/2$, learner L will with probability at least $(1 - \delta)$ output a hypothesis $h \in H$ such that $error_D(h) \leq \epsilon$, in time that is polynomial in $1/\epsilon$, $1/\delta$, n, and $size(c)$.

Vapnik-Chervonenkis dimension

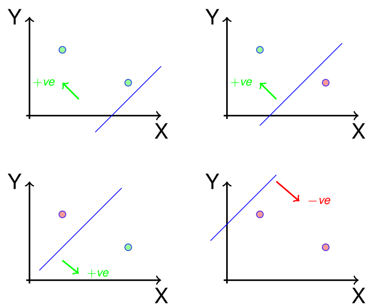
Measures the power of learner.

- The Vapnik-Chervonenkis dimension, $VC(H)$, of a hypothesis space H defined over instance space X is size of the largest finite subset of X shattered by H
- **Shattering:** a model can shatter m points $x^{(1)}, x^{(2)}, ..., x^{(m)}$ if and only if for all possible training set of $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), ..., (x^{(m)}, y^{(m)})$ there exists some parameter w such that model f_w achieves zero training error.
- Shattering can be considered as a two player game
  - Player-1 proposes d points (wish to select maximum value for d)
  - Player-2 assigns them labels
  - Player-1 provides parameters to the learner
- Consider the model $f_w(x) = sign(w^T x) = sign(w_0 x_0 + w_1 x_1 + w_2 x_2 + ...)$
- Consider the model $f_w(x) = sign(x^T x - w) = sign(\sum x_i^2 - w)$

Consider $f_w(x) = sign(w^Tx)$

Consider the model
$f_w(x) = sign(w^Tx) = sign(w_0x_0 + w_1x_1 + w_2x_2 + ...)$

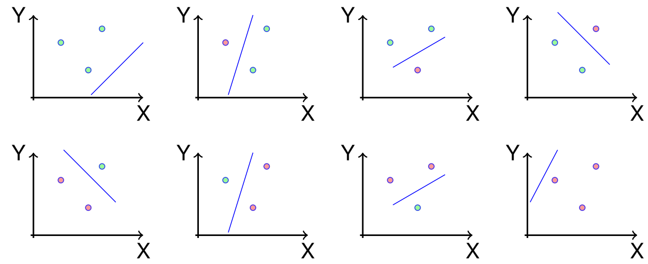
Shattering of two points



Consider $f_w(x) = sign(w^Tx)$

Consider the model
$f_w(x) = sign(w^Tx) = sign(w_0x_0 + w_1x_1 + w_2x_2 + ...)$

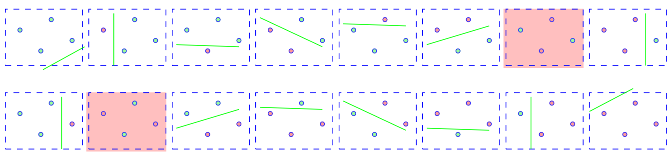
Shattering of three points



Consider $f_w(x) = sign(w^Tx)$

Consider the model
$f_w(x) = sign(w^Tx) = sign(w_0x_0 + w_1x_1 + w_2x_2 + ...)$

Shattering of four points

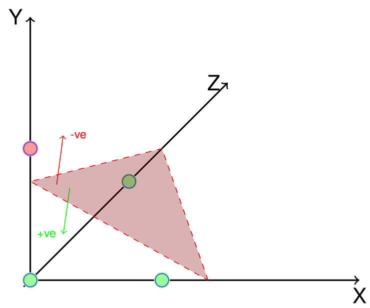


- There are cases where linear parameters are NOT provides. So four points cannot be shattered.

VC dimension of Liner classifier in d dimension

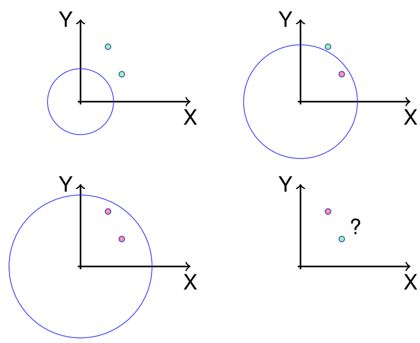
Liner classifier in d dimension have VC dimension at least $d + 1$

- d points are placed at axes and one at origin
- If label of point at origin and axes differ pass the plane through middle
- Otherwise through away from the axis point
- Shattering of d+1 point is always possible



Consider $f_w(x) = sign(x^Tx - w)$

Consider the model $f_w(x) = sign(x^Tx - w) = sign(\sum x_i^2 - w)$



Vapnik-Chervonenkis dimension

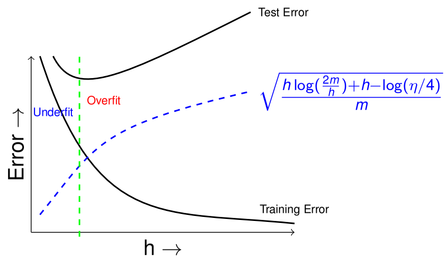
If VC dimension is h then $\exists\{x^{(1)}, x^{(2)}, ..., x^{(h)}\}$ such that $\forall\{y^{(1)}, y^{(2)}, ..., y^{(h)}\}$ $\exists w$ such that $\forall i \quad f_w(x^{(i)}) = y^{(i)}$

- VC dimension of $sign(x^Tx - w)$ is 1
- VC dimension of $sign(w^Tx)$ is 3
  Liner classifier in d dimension can have VC dimension $d + 1$
- 2^d distinct hypothesis are required to shatter d instances. Hence, $2^d \leq |H|$ therefore, $d = \boxed{VC(H) \leq \log_2 |H|}$
- Training/Empirical error $R^{train}(w) = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{2}(|y^{(i)} - f_w(x^{(i)})|)$
- Test error $R^{test}(w) = E[\frac{1}{2}|y - f_w(x)|]$
- Vapnik showed, with probability $(1 - \eta)$

$$R^{test}(w) \leq R^{train}(w) + \sqrt{\frac{h\log(2m/h) + h - \log(\eta/4)}{m}}$$

Vapnik-Chervonenkis dimension

- Consider a neural network with w number of free parameters. Then VC dimension of the neural network is
  - $O(w \log w)$ when neurons have heavy side activation $sign(v)$
  - $O(w^2)$ when neurons have sigmoid activation $sign(1/(1+e^{-v}))$



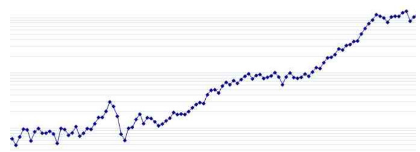
- Number of training examples required is polynomial in VC dimension

Monte Carlo Simulation

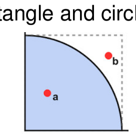
Monte Carlo simulation
uses random numbers to simulate the environment

- Stock market:

$$v_{t+1} = v_t \times a^{randomNo}$$



- Find value of π using rectangle and circle



Thank You!

Thank you very much for your attention! (Reference[1])

Queries ?

[1] [1] Ch-04/07 of the Text Book: *Machine Learning*, by *Tom M Mitchell* [2] How Self Organizing Maps (SOM) algorithm works https://www.youtube.com/watch?v=H9H6s-x-0YE [3] Lec-35 Introduction to Self Organizing Maps https://www.youtube.com/watch?v=LjJeT7rwvF4