EE4305 Fuzzy/Neural Systems for Intelligent Robotics

Owen Leong
owenleong@u.nus.edu

October 7, 2023

1 Neural Networks

1.1 Activation Functions

Step function/Threshold function/Hard limiter

$$\Phi(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Piecewise-linear neuron

$$\Phi(x) = \begin{cases} 1 & 0.5 \leq x \\ x + 0.5 & -0.5 < x < 0.5 \\ 0 & x \leq -0.5 \end{cases}$$

Logistic/Sigmoid function, with slope parameter a

$$\Phi(x) = \frac{1}{1+e^{-ax}},\ \Phi'(x) = a(1-\Phi(x))\Phi(x)$$

Tanh range $(-1, 1)$, used for hidden layers

$$\Phi(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

1.2 Neuron Models

Induced local field is

$$v_k = \sum_{i=1}^{m} w_{ki} x_i + b_k$$

If bias is included as input $x_0 = 1$ and $w_{k0} = b_k$, then

$$v_k = \sum_{i=0}^{m} w_{ki} x_i,\ y_k = \Phi(v_k)$$

1.3 Perceptron

$$v(n) = \sum_{i=0}^{m} w_i(n) x_i(n) = w^T(n) x(n)$$

$$v(n) > 0 \Rightarrow y(n) = 1,\ v(n) < 0 \Rightarrow y(n) = 0$$

- Decision boundary is the surface $w^T(n)x(n) = 0$
- For one input, decision boundary is a point, for two inputs, decision boundary is a line, for three inputs, decision boundary is a plane.
- Samples must be linearly separable for perceptron to work.
- Learning rule $w' = w + \Delta w = w + \eta e x,\ e = d - y$

1.4 Multilayer Perceptron (MLP)

Cost function

$$E(n) = \frac{1}{2}\sum_{i=1}^{m} e_i(n)^2 = \frac{1}{2}\sum_{i=1}^{m} (d_i(n) - y_i(n))^2$$

Update rule

$$w_{ji}^{(s)}(n+1) = w_{ji}^{(s)}(n) - \eta \frac{\partial E(n)}{\partial w_{ji}^{(s)}(n)}$$

s is the network layer, η learning rate

$$w_{ji}^{(s)}(n+1) = w_{ji}^{(s)}(n) + \eta \delta_j^{(s)}(n) x_{out,i}^{(s-1)}(n)$$

For output layer,

$$\delta_j^{(s)}(n) = \left(d(n) - x_{out,j}^{(s)}(n)\right) \Phi^{(s)\prime}\left(v_j^{(s)}(n)\right)$$

For hidden layer,

$$\delta_j^{(s)}(n) = \left(\sum_{k=1}^{n_{s+1}} \delta_k^{(s+1)}(n) w_{kj}^{(s+1)}(n)\right) \Phi^{(s)\prime}\left(v_j^{(s)}(n)\right)$$

Back-propagation consists of forward pass and backward pass to compute δ.

1.5 Training

Batch training vs sequential training
Normalizing inputs e.g. using mean and variance

1.6 Geometrical interpretation

- Combining basic units such as sigmoids
- Width of building block $\frac{1}{w_i^{(1)}}$
- Height of building block $w_i^{(2)}$
- Slope of building block $w_i^{(1)} w_i^{(2)}$
- Offset $b^{(2)}$, Position $b^{(1)}$

1.7 Avoiding over-fitting

1.7.1 Regularization

$$F = E_D + \lambda E_w$$

$$E_w = \sum_{i=1}^{N}\left(\left(w_i^{(1)}\right)^2 + \left(w_i^{(2)}\right)^2 + \left(b_i^{(1)}\right)^2\right) + \left(b^{(2)}\right)^2$$

$$E_w = \sum_{i=1}^{N}\left(w_i^{(2)} w_i^{(1)}\right)^2$$

1.7.2 Minimal Structure

- Using Singular Value Decomposition
- Effective rank (not numerical rank) gives insight into linear dependency in weights

2 Fuzzy Systems

2.1 Fuzzy Sets

Fuzzy set A for a discrete and finite universe of discourse X is

$$A = \sum_{i=1}^{n} \mu_A(x_i)/x_i,\ A = \int_x \mu_A(x)/x$$

where μ represents the membership and x_i is the element of discourse

$$\text{Height} = hgt(A) = \sup\{\mu_A(x)\}$$

$$\text{Support} = Supp(A) = \{x \in X : \mu_A(x) > 0\}$$

$$\text{Core} = Core(A) = \{x \in X : \mu_A(x) = 1\}$$

$$B \subseteq A \iff \mu_B(x) \leq \mu_A(x)$$

$$\text{Scalar Cardinality} = SC(A) = \sum_x \mu_A(x)$$

$$\text{Relative Cardinality} = RC(A) = \frac{SC(A)}{|X|}$$

2.2 Operations on Fuzzy Sets

2.2.1 Fuzzy Complement

- $c(0) = 1$ and $c(1) = 0$
- If $\mu_A(x) \leq \mu_A(y)$ then $c(\mu_A(x)) \geq c(\mu_A(y))$
- c is continuous
- c is involutive $c(c(a)) = a$ for $a \in [0, 1]$
- e.g. Sugeno class:

$$c_\lambda(a) = \frac{1-a}{1+\lambda a} \text{ for } \lambda \in (-1, \infty)$$

- Equilibrium of c is a where $c(a) = a$
- Common $c(a) = 1 - a$

2.2.2 Fuzzy Union (S-norm)

$$\mu_{A \cup B} = \max\{\mu_A(x), \mu_B(x)\}$$

2.2.3 Fuzzy Intersection (T-norm)

$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}$$

2.2.4 α-cut

$$A_\alpha = \{x \in X : \mu_A(x) \geq \alpha\}$$

$$A = \bigcup_{\alpha \in [0,1]} \alpha \cdot A_\alpha$$

2.3 Fuzzy Relations

Relation is defined as fuzzy set of tuples
Given fuzzy relation R and A(x)

$$\mu_B(y) = \max_{x \in X}\Big\{\min\{\mu_A(x), \mu_R(x, y)\}\Big\}$$

$$R(X, Y) \circ S(Y, Z) = \max_{y \in Y}\Big\{\min\{\mu_R(x, y), \mu_S(y, z)\}\Big\}$$

2.4 Fuzzy Functions (Extension Principle)

Given $f : X \to Y,\ y = f(x)$ then $B = f(A)$ is a fuzzy set defined by

$$\mu_B(y) = \max_{x | y = f(x)} \{\mu_A(x)\}$$

2.5 Fuzzy Number

A fuzzy number is a fuzzy set A such that A is normal, all the α-cuts of A must be closed intervals and A has a bounded support.

2.5.1 Operations on Fuzzy Numbers (e.g. $*$)

$$(A * B)(z) = \max_{z = x * y} \Big\{\min\{A(x), B(y)\}\Big\}$$

Alternatively,

$$A * B = \bigcup_{\alpha \in [0,1]} \alpha \cdot (A * B)_\alpha$$

2.5.2 Operations on Intervals

$$[a, b] + [c, d] = [a + c, b + d]$$

$$[a, b] - [c, d] = [a - d, b - c]$$

$$[a, b] * [c, d] = \Big[\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)\Big]$$

$$\frac{[a, b]}{[c, d]} = \Big[\min(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}), \max(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d})\Big], 0 \notin [c, d]$$

2.6 Fuzzy Inference

Mamdani Implication $\delta(a, b) = \min\{a, b\}$

2.6.1 Set up

Rule - If X is A then Y is B
Fact - X is A'
Conclusion - Y is B'

2.6.2 Steps

Step 1: Degree of fulfilment

$$r(A') = hgt(A' \cap A)$$

Step 2: Draw conclusion

$$B'(y) = \min(r(A'), B(y))$$

For multiple rules, take the union

$$B'(y) = \bigcup_k B'_k(y)$$

2.7 Fuzzy Knowledge Based Control (FKBC)

2.7.1 Fuzzification

Convert input into fuzzy variable.

2.7.2 Inference

Using if-then rules as described previously.

2.7.3 Defuzzification

Center-of-mass

$$x^* = \frac{\int \mu_A(x) x \, dx}{\int \mu_A(x) \, dx} - \frac{\sum x_i \mu_A(x_i)}{\sum \mu_A(x_i)}$$

3 Extras

3.1 Vector Derivatives

$$\frac{d(x^T B)}{dx} = B, \ \frac{d(x^T b)}{dx} = b$$

$$\frac{d(x^T x)}{dx} = 2x, \ \frac{d(x^T A x)}{dx} = 2Ax$$