Terms Used in the Lectures
(in alphabetic order)

α-cut (pronounced *alpha-cut*). Let:

- $\mu(x)$ be a membership function (see), and
- α be a positive number from the interval $[0,1]$.

The α-cut of $\mu(x)$ is the set $\{x : \mu(x) \ge \alpha\}$ of all the values x for which the degree $\mu(x)$ is greater than or equal to α.

Activation function. In a neural network, signals interchangingly undergo:

- linear transformations and
- non-linear transformations.

The corresponding nonlinear transformation is known as an activation function.

Algebraic product. The usual product $a \cdot b$ of two numbers.

"And"-operation. An algorithm that transforms:

- our degrees of confidence a and b in statements A and B
- into an estimate $f_\&(a,b)$ for our degree of confidence in the statement

$$\text{“}A \text{ and } B\text{”} \ (A \,\&\, B).$$

Averaging. It is when we replace:

- several values $a, \ldots, a_n$ with
- their arithmetic average $\dfrac{a_1 + \ldots + a_n}{n}$.

Centroid defuzzification. A defuzzification (see) in which:

- based on a membership function $\mu(x)$,
- we generate the value $\overline{x} = \dfrac{\int x \cdot \mu(x)\,dx}{\int \mu(x)\,dx}$.

Deep learning. Machine learning (see) that is based on a deep neural network (see).

Deep neural network. In the most general understanding, data processing in which a signal goes through many different stages.

Defuzzification. An algorithm that transforms:

- a membership function $\mu(x)$ (see)
- into a single number $\overline{x}$.

Explainable AI. It is when:

- an AI system not only provides *recommendations*,
- it also provides *explanations* for these recommendations.

Fuzzy set. Same as membership function (see).

Fuzzy methodology. Same as fuzzy techniques (see).

Fuzzy techniques. Techniques for translating:

- expert knowledge which has been formulated by using imprecise ("fuzzy") words from a natural language (like "small")
- into precise computer-understandable terms.

Interpolation/extrapolation.

- In several cases $k = 1, \ldots, K$, we know the values $x^{(k)}$ and $y^{(k)}$ of quantities x and y.
- We want to find an algorithm $f(x)$ which fits all this data, i.e., for which, for all k, we have $f\left(x^{(k)}\right) \approx y^{(k)}$.

K-Lischitz function. A function $f(x)$ is called a K-Lipschitz function if it satisfies Lipschitz condition (see) with parameter K.

Least Squares approach. In many practical situations:

- it is reasonable to assume that the dependence of y on x is determined by an expression $f(x, c)$ for some parameters c; and
- we know the values $x^{(k)}$ and $y^{(k)}$ corresponding to several cases $k = 1, \ldots, K$.

The least squares approach is to select the values c for which the following sum $\sum\limits_{k=1}^{K}\left(y^{(k)}-f\left(x^{(k)},c\right)\right)^2$ is the smallest possible.

Linear interpolation. An interpolation (see) in which the corresponding function $f(x)$ is linear.

Lischitz condition. A function $f(x)$ satisfies Lipschitz condition with parameter K if for all x and x', we have $|f(x)-f(x')|\le K\cdot|x-x'|$.

Logistic activation function. Same as sigmoid activation function (see).

Machine learning. Same as interpolation/extrapolation (see).

Max-pooling. It is when we replace:

- several values $a_1,\ldots,a_n$
- with a single value $\max(a_1,\ldots,a_n)$.

Maximally (most) individually robust function $f(x)$ of one variable. Let us assume that we are given a class of functions. A function from this class is called maximally (most) individually robust if, for all x and x', it satisfies the inequality $|f(x)-f(x')|\le K\cdot|x-x'|$ with the smallest possible value K.

Maximally (most) individually robust function $f(x,y)$ of two variables. Let us assume that we are given a class of functions. A function from this class is called maximally (most) individually robust if, for all x, y, x', and y', it satisfies the inequality $|f(x,y)-f(x',y')|\le K\cdot\max(|x-x'|,|y-y'|)$ with the smallest possible value K.

Membership function. For each natural-language expression P (e.g., "small"), a membership function is a function that assigns:

- to each possible value x of a quantity,
- the degree $\mu_P(x)$ to which this value satisfies the given property (e.g., to which this value is small).

Negation operation. An algorithm that transforms:

- our degree of confidence a in a statements A
- into an estimate $f_\neg(a)$ for our degree of confidence in the statement

$$\text{“not } A\text{” } (\neg A).$$

"Or"-operation. An algorithm that transforms:

- our degrees of confidence a and b in statements A and B
- into an estimate $f_\vee(a,b)$ for our degree of confidence in the statement

$$\text{“}A \text{ or } B\text{”}\ (A \vee B).$$

Polling. A method for finding a degree of confidence in a statement S:

- we ask N experts, and
- if M of them thing that this statement is true, we take M/N as the desired degree.

Pooling. It is when we replace:

- several values $a_1, \ldots, a_n$
- with a single value.

Rectified linear activation function. A function $F(x) = \max(0, x)$.

Robust. A function $f(x)$ is called robust if:

- whenever x is close to x' (denoted by $x \approx x'$),
- the values $f(x)$ and $f(x')$ of the function are also close: $f(x) \approx f(x')$.

Sigmoid activation function. A function $F(x) = \dfrac{1}{1 + \exp(-x)}$.

Sum-pooling. It is when we replace:

- several values $a, \ldots, a_n$ with
- their sum $a_1 + \ldots + a_n$.

t-Conorm. Same as "or"-operation (see).

t-Norm. Same as "and"-operation (see).

Tuning. We have:

- the data $\left(x^{(k)}, y^{(k)}\right)$, and
- an algorithm $f_0(x)$ that fits the data, e.g., for which $f_0\left(x^{(k)}\right) \approx y^{(k)}$.

Tuning means finding an algorithm $f(x)$ that provides a better fit with the data.

Variational derivative. A derivative with respect to a variable which is a value of the unknown function.

Variational optimization. An optimization problem in which the unknown is a function.