MACHINE LEARNING

26 November 2021

---

1 Checkbox questions

1. (3 p.) Which one of the following are regression problems?

   ☐ Decide when to buy and when to sell on the stock market on the basis of a window of previous samples;

   ☐ Decide whether two fingerprints belong to the same person;

   ☐ Predict the annual income of a company on the basis of the field of business and on the number of employees;

   ☐ Nothing of the above.

y x ~ 2 ~ 50

2. (3 p.) What is the meaning of overfitting?

   ☐ It is a synonym of "best fitting";

   ☐ It is refers specifically to the LMS algorithm, for the case of quadratic loss;

   ☐ It indicates a fitting of the training set with scarse degree of parsimony;

   ☐ Nothing of the above.

3. (3 p.) Which one of the following is correct concerning the saturation of sigmoidal neurons?

   ☐ Sigmoidal neurons saturates when the value of the weights become big;

   ☐ Sigmoidal neurons never saturates;

   ☐ The saturation of sigmoidal neurons is independent of the input.

   ☐ Nothing of the above.

4. (3 p.) Let us consider the supposed loss function

$(0,1)$

$$V(f(x), y) = -y \log f - (1-y)\log(1-f)$$

where $y \in \{0,1\}$ is the target and f is the value returned by a sigmoidal neuron in the scalar case. Which of the following holds true?

   ☐ This is an entropy, but it is not a loss function since it returns negative values;

   ☐ This loss function is typically better than the quadratic loss for classification;

   ☐ The above entropy loss can also be used with targets in $\{-1,+1\}$;

   ☐ Nothing of the above.

$(-1,+1)$

$-\frac{1+z}{2}\log\frac{1+f}{2} - \left(1-\frac{1+z}{2}\right)\log\left(1-\frac{1+f}{2}\right)$

$-\frac{1+z}{2}\log\frac{1+f}{2} - \frac{1-z}{2}\log\frac{1-f}{2}$

5. (3 p.) Let us consider the empirical risk function

$$E = \left(\sum_{\kappa=1}^{\ell} (y_\kappa - f(w, x_\kappa))^{2m}\right)^{1/2m}$$

where $m \in \mathbb{N}$. Which of the following statements is true?

$m = 1$

$(e_1^{2m} + e_2^{2m} + e_3^{2m})^{1/2m}$

$(e_1^2 + e_2^2 + e_3^2)^{1/2}$

$(e_1^{200} + e_2^{200} + e_3^{200})^{1/200}$

- ☐ The learning with this empirical risk always returns a perfect match $E \to 0$ for $m \to \infty$;
- ☐ This empirical risk returns the maximum error $\max_\kappa |(y_\kappa - f(w, x_\kappa)|$ independently of learning as $m \to \infty$; $k = 1, .., \ell$
- ☐ This empirical risk returns the maximum error $\max_\kappa |(y_\kappa - f(w, x_\kappa)|$ as $m \to \infty$ only at the end of learning;
- ☐ Nothing of the above.

6. (3 p.) Which of the following statements is true concerning the regularization parameter in ridge regression? λ
   - ☐ The regularization parameter can be any small real number; $\lambda \in \mathbb{R}$
   - ☐ The regularization parameter leads to discover a unique solution in normal equations;
   - ☐ The regularization parameter improves the fitting on the training set;
   - ☐ There is always a unique solution in normal equations also if the regularization parameter is zero.
   - ☐ Nothing of the above.

$R + \lambda ws^2$

2 Open questions

1. (4 p) Design a neural network — by explicitly ~~indicating~~ neurons the values of the weights — that is composed of the cascade of two rectifiers with the purpose of realizing the function shown in Fig. 1

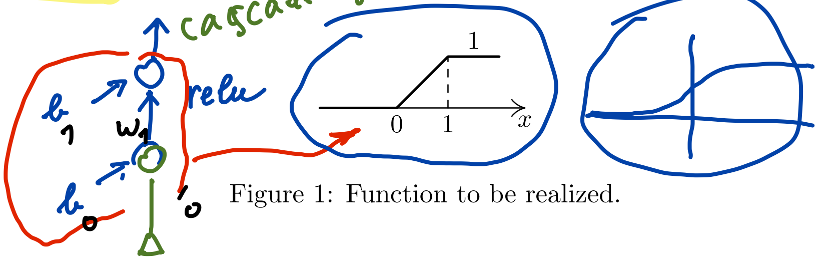
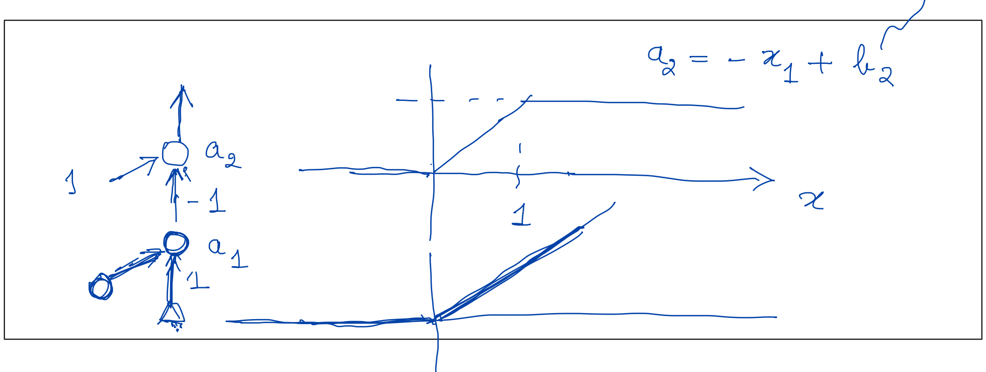


Figure 1: Function to be realized.

$\frac{1}{1+e^{-a}} = \sigma(a)$

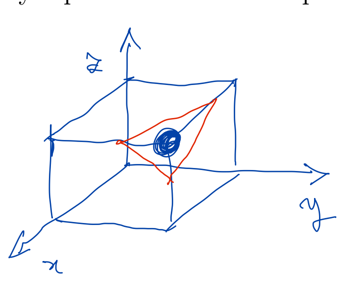


2. (4 p) Consider the Boolean function

$$f(x, y, z) = x \wedge y \wedge z.$$

Is it linearly-separable? Proof is required.

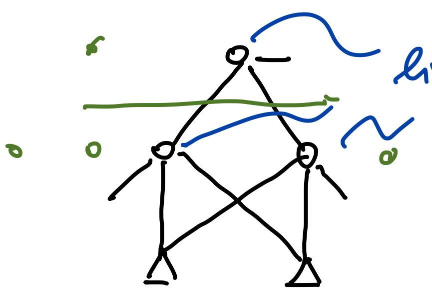
x	y	z	∧
	⋮		



3. (4 p) Suppose you are given a multilayered neural networks with two inputs, one output, and any number p of arbitrarily large hidden layers. If the neurons are linear, that is

$$x = \sigma(a) = a,$$

can this neural network compute the XOR predicate? Motivate the answer.



linear

$x = a$

$a_i = \sum_j w_{ij} a_j$

4. (Optional. 6 p) Consider a collection of black & white pictures and suppose we want to separate those with more black than white pixels. Can we solve this problem by a neural network with one sigmoidal neuron only? Proof is required.