Introduction to Machine Learning Course

Dry 6 – Deep learning

Submitted individually by Thursday, 01.07, at 23:59. Each day of delay costs 5 points.

You may answer in Hebrew or English and write on a computer or by hand (but be clear).

Please submit a PDF file named like your ID number, e.g., *123456789.pdf*.

Bonus (maximal grade is 100): Writing on a computer (using LyX/LaTeX, Word + Equation tool, etc.) = 3 pts.

1. Consider a trained fully connected neural network with L linear layers.

   Denote the function of the network by $F_\Theta: \mathbb{R}^d \to \mathbb{R}$, where $\Theta = \left(\underbrace{\mathbf{W}^{(1)}}_{\in \mathbb{R}^{d\times p}}, \underbrace{\mathbf{W}^{(2)}}_{\in \mathbb{R}^{p\times p}}, \ldots, \underbrace{\mathbf{W}^{(L-1)}}_{\in \mathbb{R}^{p\times p}}, \underbrace{\boldsymbol{w}^{(L)}}_{\in \mathbb{R}^{p}} \right)$ is the set of all weights and no biases (we follow the notations from Tutorial 11).

   As an activation function, we use the ReLU function $\sigma(z) = \max\{0, z\}$.

   The network's output is given by:

   $$F_\Theta(x) = \boldsymbol{w}^{(L)^\top} \boldsymbol{h}^{(L-1)}(x)$$

   where we recursively define the hidden layers:

   $$\boldsymbol{h}^{(1)}(x) = \sigma\left(\mathbf{W}^{(1)^\top} x\right), \qquad h^{(\ell)}(x) = \sigma\left(\mathbf{W}^{(\ell)^\top} \boldsymbol{h}^{(\ell-1)}(x)\right)$$

   We now scale all weights in Θ by a factor of $\alpha \in \mathbb{R}_{>0}$.

   Notice: the ReLU function is positive-homogeneous in the sense that $\sigma(\alpha \cdot z) = \alpha \cdot \sigma(z)$.

   1.1. Show that the new output function holds $F_{\alpha\cdot\Theta}(x) = c \cdot F_\Theta(x)$ for some scalar $c \in \mathbb{R}$.

   In your answer, find an appropriate c value for which this statement holds.

   We wish the model to output a probability. As seen in Lecture 09 (for logistic regression), we can apply the sigmoid function to F_Θ. That is, the model's output will be: $\frac{1}{1+\exp\{-F_{\alpha\cdot\Theta}(x)\}}$.

   1.2. For $\alpha \to 0$, to which probability does the output converge?

   1.3. For $\alpha \to \infty$, to which probability does the output converge?