Review Questions 1

Group 1
Chuan Su
Diego Alonso Guillen Rosaperez

November 16, 2019

1. **a**) and **c**) about *Random Forest*.
   (a) Individual tree is built on a subset of the features.
   (c) Individual tree is built on a subset of instances.

2. Number of features to use as candidates for splitting at each tree node. The number is specified as a fraction or function of the total number of features. Decreasing this number will speed up training, but can sometimes impact performance if too low.

3. If the instances are only one class, there is only one potential label and, therefore, the probability of getting that label is 1. In that sense:

$$entropy(D) = -1 * log_2(p_i = 1) = 0$$

4. Gini index measures the misclassification (also named impurity). Since in a pure partition there exists only one class, the probability of obtaining that class is 1. In that sense:

$$Gini(D) = 1 - (p_i = 1)^2 = 0$$

5. Given the hidden function $h(x) = w_1^\intercal x$ and out function $g(x) = w_2^\intercal x$. The feedforward neural network function of input x is

$$f(x) = g(h(x)) = w_2^\intercal(w_1^\intercal x) = (w_2^\intercal w_1^\intercal)x$$

   $w_2^\intercal w_1^\intercal$ is a matrix multiplication and results in an linear function of input x.

6. On one hand, back-propagation is run to train multilayer neural network. Back-propagation requires differentiable activation function and uses its derivative as a multiplier to update weights. However derivative of step function is **0**, which implies that gradient descent won't be able to make progress in updating the weight. On the other hand, we want that the prediction be as close as the real value. To optimize the weights to achieve this, we want that a small change in the input result in a small change on the output, not that it changes only between zero and one, which would not help us to tweak the weights.

7. **Forward pass**

$$net_{h2} = w_3 x_1 + w_4 x_2 + b_1 = 0.25 \times 0.05 + 0.30 \times 0.1 + 0.35 = 0.3925$$
$$net_{h1} = w_1 x_1 + w_2 x_2 + b_1 = 0.15 \times 0.05 + 0.2 \times 0.1 + 0.35 = 0.3775$$

$$out_{h1} = max(0, net_{h1}) = net_{h1} = 0.3775$$
$$out_{h2} = max(0, net_{h2}) = net_{h2} = 0.3925$$

$$net_{o1} = w_5 out_{h1} + w_6 out_{h2} + b_2 = 0.4 \times 0.3775 + 0.45 \times 0.3925 + 0.6 = 0.9276$$
$$net_{o2} = w_7 out_{h1} + w_8 out_{h2} + b_2 = 0.5 \times 0.3775 + 0.55 \times 0.3925 + 0.6 = 1.0046$$

$$out_{o1} = max(0, net_{o1}) = 0.9276$$
$$out_{o2} = max(0, net_{o2}) = 1.0046$$

$$E_{o1} = \frac{1}{2}(target_{o1} - out_{o1})^2 = \frac{1}{2}(0.01 - 0.9276)^2 = 0.42099$$
$$E_{o2} = \frac{1}{2}(target_{o2} - out_{o2})^2 = \frac{1}{2}(0.99 - 1.0046)^2 = 0.000106$$
$$E_{total} = E_{o1} + E_{o2} = 0.421096$$

Backward pass $\mathbf{w_8^{(next)}}$

$$\frac{\partial E_{total}}{\partial w_8} = \frac{\partial E_{total}}{\partial out_{o2}} \times \frac{\partial out_{o2}}{\partial net_{o2}} \times \frac{\partial net_{o2}}{\partial w_8}$$

$$\frac{\partial E_{total}}{\partial out_{o2}} = -(0.99 - 1.0046) = 0.0146$$
$$\frac{\partial out_{o2}}{\partial net_{o2}} = \begin{cases} 0 & net_{o2} < 0 \\ 1 & net_{o2} > 0 \end{cases}$$
$$\frac{\partial net_{o2}}{\partial w_8} = out_{h2} = 0.3925$$
$$\frac{\partial E_{total}}{\partial w_8} = 0.0146 \times 1 \times 0.3925 = 0.0057305$$
$$w_8^{(next)} = w_8 - \eta \frac{\partial E_{total}}{\partial w_8} = 0.55 - 0.5 \times 0.0057305 = 0.54713$$

Backward pass $\mathbf{w_2^{(next)}}$

$$\frac{\partial E_{total}}{\partial w_2} = \frac{\partial E_{total}}{\partial out_{h1}} \times \frac{\partial out_{h1}}{\partial net_{h1}} \times \frac{\partial net_{h1}}{\partial w_2}$$

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$

$$\frac{\partial \mathbf{E_{o1}}}{\partial \mathbf{out_{h1}}} = \frac{\partial E_{o1}}{\partial out_{o1}} \times \frac{\partial out_{o1}}{\partial net_{o1}} \times \frac{\partial net_{o1}}{\partial out_{h1}}$$

$$\frac{\partial E_{o1}}{\partial out_{o1}} = -(target_{o1} - out_{o1}) = -(0.01 - 0.9276) = 0.9176$$

$$\frac{\partial out_{o1}}{\partial net_{o1}} = \begin{cases} 0 & net_{o1} < 0 \\ 1 & net_{o1} > 0 \end{cases}$$

$$\frac{\partial net_{o1}}{\partial out_{h1}} = w_5 = 0.40$$

$$\frac{\partial E_{o1}}{\partial out_{h1}} = 0.9176 \times 1 \times 0.40 = 0.36704$$

$$\frac{\partial \mathbf{E_{o2}}}{\partial \mathbf{out_{h1}}} = \frac{\partial E_{o2}}{\partial out_{o2}} \times \frac{\partial out_{o2}}{\partial net_{o2}} \times \frac{\partial net_{o2}}{\partial out_{h1}}$$

$$\frac{\partial E_{o2}}{\partial out_{o2}} = -(target_{o2} - out_{o2}) = 0.0146$$

$$\frac{\partial out_{o2}}{\partial net_{o2}} = \begin{cases} 0 & net_{o2} < 0 \\ 1 & net_{o2} > 0 \end{cases}$$

$$\frac{\partial net_{o2}}{\partial out_{h1}} = w_7 = 0.5$$

$$\frac{\partial E_{o2}}{\partial out_{h1}} = 0.0146 \times 1 \times 0.5 = 0.0073$$

$$\frac{\partial \mathbf{E_{total}}}{\partial \mathbf{out_{h1}}} = 0.36704 + 0.0073 = 0.37434$$

$$\frac{\partial out_{h1}}{\partial net_{h1}} = \begin{cases} 0 & net_{h1} < 0 \\ 1 & net_{h1} > 0 \end{cases} = 1$$

$$\frac{\partial net_{h1}}{\partial w_2} = x_2 = 0.10$$

$$\frac{\partial E_{total}}{\partial w_2} = 0.37434 \times 1 \times 0.10 = 0.037434$$

$$w_2^{(next)} = w_2 - \eta \frac{\partial E_{total}}{\partial w_2} = 0.20 - 0.5 \times 0.037434 = 0.181283$$