1st progress report

Mamoru Shibata

2020/3/19

1 Describe the 4 encoding methods

Define a set of input values as $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_p})$. p is the types of features. And $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix}$, $x_{ij} \in \mathcal{X}_i$, $|\mathcal{X}_i| = N_i$. The target feature is $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $y_i \in (0, 1)$.

1.1 One-Hot-Coding

Define The function of One-Hot Coding as $OHC(\mathbf{x})$.

$$OHC(\mathbf{x}_i) = (e^i_1, e^i_2, ..., e^i_{N_i}) \tag{1}$$

$$e^i_1 + e^i_2 +, ..., + e^i_{N_i} = 1 \tag{2}$$

$$e^i_j = \begin{cases} 1 & (x_{ij} = e^i_j) \\ 0 & (otherwise) \end{cases} \tag{3}$$

1.2 Label Encoding

Define The function of Label encoding as $LE(\mathbf{x})$.

$$LE(\mathbf{x}_i) = \begin{pmatrix} t_{i1} \\ t_{i2} \\ \vdots \\ t_{in} \end{pmatrix} \tag{4}$$

$$t_{ij} \in \mathbb{Z} \tag{5}$$

$$t_{ij} = t_{ik} \text{ only if } x_{ij} = x_{ik} \tag{6}$$

1.3 Target Encoding

Define The function of Label encoding as $TE(\mathbf{x})$.

$$TE(\mathbf{x}_i) = \begin{pmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{in} \end{pmatrix} \tag{7}$$

$$s_{ij} \in \mathbb{R} \tag{8}$$

$$s_{ij} = \frac{\sum_{k, x_{ik} = x_{ij}, y_k = 1}}{\sum_{k, x_{ik} = x_{ij}}} \tag{9}$$

1.4 One-Hot-Coding + PCA

Define The function of OHC+PCA as $OP(\mathbf{x})$.

$$OHC(\mathbf{x}_i) = (e^i_1, e^i_2, ..., e^i_{N_i}) \tag{10}$$

$$OP(\mathbf{x}_i) = (e'^1_1, e'^i_2, ..., e'^i_{N'_i}) \tag{11}$$

$$N_i > N'_i \tag{12}$$

2 Describe the 3 learning methods

2.1 SVM

1. C : The penalty of misclassification. This parameter is decided by Cross Validation.
2. Kernel : The kernel function is "RBF". This function give the weight for points. $K(x, x') = \exp(\frac{\gamma ||x - x'||^2}{2\sigma^2})$
3. Probability : True. The predict probability is "The probability of collect prediction"???
4. gamma : The parameter of kernel. Use "scale". So, $\gamma = 1/(\text{n\_features} * \text{X\_variance})$

2.2 Decision Tree

The all parameters are default.

1. criterion : "gini". The function to measure the quality of a split.
2. splitter : "best". The strategy used to choose the split at each node.
3. max_depth : "None". The maximum depth of the tree.
4. min_samples_leaf : "1". The minimum number of samples required to be at a leaf node.
5. max_iter : "3000". Maximum number of iterations in the solver.

2.3 Neural Network

1. hidden_layer_sizes : "(100,)". The ith element represents the number of neurons in the ith hidden layer. So, the number of hidden layers is "1".
2. activation : "ReLU". Activation function for the hidden layer. $f(x) = \max(0, x)$