University of Tehran
Faculty of New Sciences and Technologies Department
of Mechatronics

Practice of Artificial Neural Networks

By:

Omid Moradi

Lecturer:

Dr. Masoudnia

Fall 2024

Q1. Explain the general and multi-purpose MRMR (Minimum Redundancy-Maximum Relevance) feature selection technique.

The MRMR (Minimum Redundancy-Maximum Relevance) method is a common approach for selecting features in machine learning, especially when dealing with high-dimensional data. The goal is to select a subset of features that are most relevant to the target variable while avoiding redundancy among the selected features. MRMR is useful in different types of tasks like classification, regression, and clustering, and is applied in fields such as bioinformatics, text analysis, and other domains that involve large amounts of data.

Maximum Relevance: This ensures the selected features have a strong relationship with the target variable. Relevance is typically measured using mutual information, which quantifies how much one variable tells us about another.

Minimum Redundancy: This focuses on minimizing the overlap between the features. In other words, it tries to select features that provide unique information, reducing correlations or redundancies among them.

How it works:

1. First, we measure the relevance of each feature with respect to the target variable, often using mutual information.
2. Then, the redundancy between the features is measured, and features that offer the least overlap (in terms of information) are chosen.
3. The final selection is a balance between maximum relevance and minimum redundancy.

```python
# create some pandas data
import pandas as pd
from sklearn.datasets import make_classification
X, y = make_classification(n_samples = 1000, n_features = 50, n_informative = 10, n_redundant = 40)
X = pd.DataFrame(X)
y = pd.Series(y)

# select top 10 features using mRMR
from mrmr import mrmr_classif
selected_features = mrmr_classif(X=X, y=y, K=10)
```

Q2. The MSE term can be written based on the Bias and Variance terms, known as the Bias-Variance decomposition of error. Explain and prove it.

n machine learning, the **Mean Squared Error (MSE)** of a model can be broken down into three parts: **Bias**, **Variance**, and an **Irreducible Error**. This is known as the Bias-Variance Decomposition and helps explain the trade-offs between model complexity and performance.

$$MSE = E\left[\left(\bar{f}(x) - f(x)\right)^2\right]$$

We can decompose the error as follows:

$$MSE = E\left[\left(\hat{f}(x) - E[\hat{f}(x)] + E[\hat{f}(x)] - f(x)\right)^2\right]$$

$$= \mathrm{E}[(\hat{f}(x) - E[\hat{f}(x)^2] + (E[\hat{f}(x)] - f(x))^2$$

$$MSE = Variance + Bias2 + Irreducible\ Error$$

The decomposition is based on expanding the error and simplifying it by separating the bias and variance components, showing that the total error is a sum of these two plus the irreducible error.