Класификация с помощта на метода k-NN

Изготвил:

Радостина Иванова ФН:45464

1. **Представяне на обучителното множество и тестовия пример**

Стойностите на оборота се съхраняват в променливи от тип double.

Стойносттите на характеристиката на екипа се съхраняват в променлива от тип double по следния начин:

Ако характеристиката е Weak, то се съхранява в променливата,чиято стойност ще е 1;

Ако характеристиката е Average, то се съхранява в променлива, чиято стойност ще е 2;

Ако характеристиката е Strong, то се съхранява в променлива, чиято стойност ще е 3;

Ако характеристиката е Very Strong, то се съхранява в променлива, чиято стойност ще е 4;

Последната - успехът се представя чрез променлива от тип double със стойност 0 или 1-ца;

Един ред от таблицата, който съдържа трите атрибута, е представен чрез vector<double>, а всичките съвкупно като vector<vector<double>>;

1. **Нормализация**

Осъществява се с функцията normalizе, която за всеки елемент го представя чрез следната формула:

**A screenshot of a cell phone

Description automatically generated**

1. **Функции за изчисляване на разстоянието**

Първата функция изпозва euclidean distance.

Втората функция използва сумата от следващите две:

a)За характеристика на екипа използва Hamming distance т.е ако съвпадат връща 0 за разстояние, ако не съвпадат 1-ца

b) За оборота разликата между стойностите на този атрибут на даден обучителен пример и тестовия.

1. **K - nearest neighbors algorithm**

Изчисляване на разстоянието от тестовия пример с обучителните като създава

pair от типа: разстояние и обучителен пример като ги съхранява

във vector<double, vector<double>>

Сортира вектора във възходящ ред спрямо разстоянието и после взима само първите K.

Намира колко от тях са категоризирани фалирали и колко от тях успешни и към тази категория, която е по-срещана, се причислява тестовия пример.

Ако са еднакъв брой се причислява към тази, категория, чийто наследник е най-малко разстояние до тестовия пример.

1. **Въвеждане на обучителните примери**

Използва csv разширение, което показва данните като текст(всяка колона е разделена с „ , “, а всеки обучителен пример е на нов ред). Текста с данните се трансформира с функциите getData и split.

1. **Въвеждане на тестови пример**

Въвежда се през конзолата.

Първо се въвежда оборота, а после за характеристика на екипа стойност

имежду Very Strong, Strong, Average и Weak.

1. **Резултат от програмата**

Извежда дали компанията е класифицирана като **Bankrupt** или **Successful**

**8.Избор на K**

K1 = sqrt(Броя на обучителните примери)

Няколко правила при избора на следващите два избора на K

1.K трябва да е нечетно число

2. К стойността не трябва да е кратна на броя класове

3.К не трябва да е прекалено голямо или прекалено малко

K2 = 9

K3 = 19