Machine Learning for Grading Students

Gautam Singh

Contents

***Abstract*—This document is a report which compares the grade distribution obtained by using a method based on machine learning as compared to fitting a normal curve to the scores of the students.**

1 Introduction

We test the utility of the K-means algorithm in assigning grades as compared to estimating the grades using the standard normal distribution.

We consider the scores of $N = 94$ students who have taken a course in the Indian Institute of Technology, Hyderabad (IITH) as our dataset.

2 Fitting a Gaussian Curve

Since N is not very large, given the scores of each student x_i, $1 \leq i \leq N$, we can compute the population mean and population variance as

$$\mu = E[x] \tag{1}$$

$$\sigma^2 = E\left[(x-\mu)^2\right] \tag{2}$$

We assume that the scores $x \sim N\left(\mu, \sigma^2\right)$. Thus, we compute the Z-scores as

$$Z = \frac{x-\mu}{\sigma} \tag{3}$$

The grades are assigned as per the following table.

Interval	**Grade**
$(-\infty$, -3]	F
(-3, -2]	D
(-2, 1]	C
(-1, 0]	B-
(0, 1]	B
(1, 2]	A-
(2, 3]	A
(3,∞)	A+

TABLE 1: Grading Scheme.