Assignment-3: Basics of Statistics

IITG AI-ML Course

March 11, 2020

1. Visualizing mean, median, mode and their weighted versions:

   (a) Generate a 1-d dataset by randomly drawing 1000 points from an uniform distribution of range[-100,100].

   (b) Generate a set of 1000 weight values by drawing 1000 values from a normal distribution of mean 0 and variance 4.

   (c) Compute the mean, median, mode.

   (d) Compute the weighted mean, weighted median, weighted mode, where the ith weight value corresonds to the ith element in the dataset.

      i. The weighted mean can be computed from the formula: $\frac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i}$

      ii. For calculation of the weighted median, the frequency value for the ith point is given by $w_i / \Sigma w_i$.

      iii. For the weighted mode, the weight w_i is proportional to the number of occurence of the ith element.

   (e) Visualize the mean, median, mode and their weighted versions on a histogram with proper colouring to distinguish them.

2. Correlation between variables:

   (a) Consider the equation of the straight line $y = mx$ where m = $tan\,\theta$ and θ varies from 5° to 185° in steps of 10°.

   (b) Add random Gaussian noise to each of the y values obtained from the straight line equation to obtain a set of noisy values $\hat{y}$

   (c) Compute the correlation co-efficient between x and $\hat{y}$ for each value of θ.

   (d) Plot the correlation values against the θ values.