Generating the dataset

General information

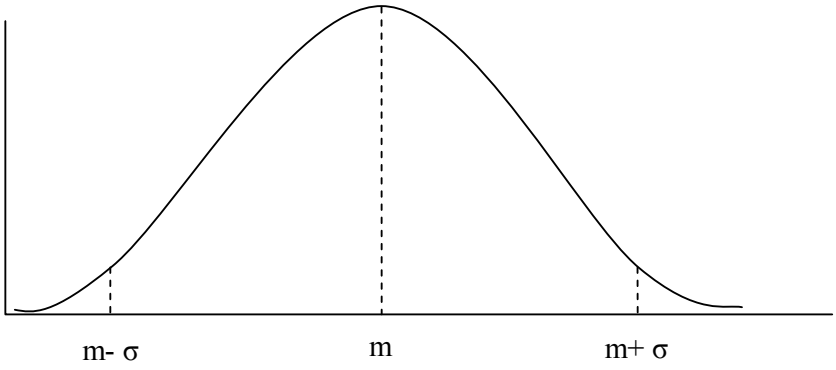
For a better analyzing of the learning algorithms results that will be developed in this laboratory, the dataset that will be used will contain a lot of points that can be plotted into a two-dimensional space. Thus, every entry (sample) from the dataset will have two features: the first represent the coordinate *x* and the second represent the coordinate *y* of a point that will be displayed on screen. In this idea I will propose an algorithm, based on Gauss function that generates these points randomly.

Gaussian probability

The Gaussian probability is the probability that a point to be into the interval $[m-\sigma, m+\sigma]$, where *m* is the center (middle range of curve) and σ is the dispersion of data.

The Gauss function:

$$Gauss(x) = e^{-\frac{(m-x)^2}{2\sigma^2}}$$



The algorithm to generate the coordinates randomly

This algorithm will be run independently for each feature of a sample from the dataset (for the coordinate *x* and after that, for the coordinate *y* of each point). The parameters *m* and σ are specified, as entries of the algorithm, for each group of data that will be created. Thus each data group that will be created will have *mx*, *my*, σ_x and σ_y own. Also each group will have a different color. The values of *x* and *y* will be represented in the Cartesian coordinates not in screen coordinates.
The algorithm:

1. It is choused randomly a group for generating one coordinate.
2. Randomly will be chooses a value for the coordinated *x* (in the *x* domain);

3. Will be calculate the probability that randomly selected value (at step 2) is or is not closer to center *m* (using the Gauss function);
4. Will be generate randomly a probability in [0,1] domain;
5. It is check, if the probability calculated for our choused coordinate (at step 3) is greater than the probability randomly choused (at step 4);
    a. If yes, the coordinated randomly choused (at step 2) will be consider and pass to step 5;
    b. If not, go back to step 2;
6. Make the steps 2-5 for the coordinated y, then in step 5.a algorithm terminates for one sample (one point).

For a center *m* and a dispersion σ, that are gives, the algorithm will generate a randomly numbers of points around *m*.

For this laboratory is required to make an application that generates approximately 3000 points for 3 different centers. The point will be writing into a file, each point per line, and also will be drawn to the screen. Will be generated the points for three different centers and dispersions. Lines from the file with points characteristics for a center it is preferred to not be consecutively.

Note:
The screen center will be considered as the axes center of coordinates (0,0). Means that points will be in Cartesian coordinates (not Screen coordinates).