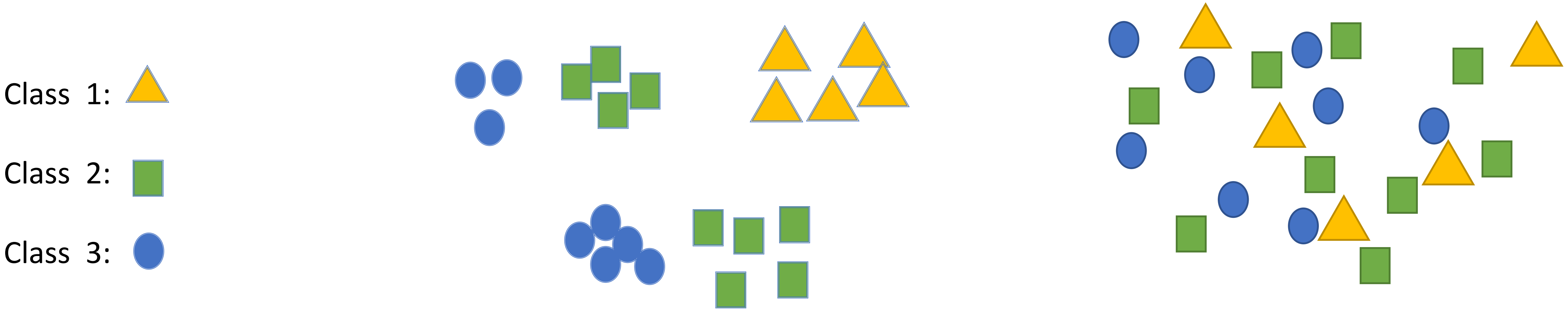
Separation and Smoothness Indices for classification and regression problems

Ahmad Kalhor

Separation Index (SI) for Classification problems

- Assume in a classification problem, a real n-dimensional vector space includes N examples that each one belongs to a certain class from m classes.
- The separation index denotes the separability among examples of different classes.
- Following figure shows two different states of separability among examples of three classes 1 ,2,3:



High Separability

Low Separability

How do we define the SI in a classification problem?

- Consider we have N examples from “m” different classes in a certain n-dimensional vector space.
- Assume N1 denotes the number all possible examples whose nearest neighbors (in Lp norm) have the same labels with them. The (first order) SI is defined as follows:

$$SI = \frac{N1}{N}, \quad 0 \leq SI \leq 1$$

$$SI = 1(perfect\ 1st\ order\ sepration)$$

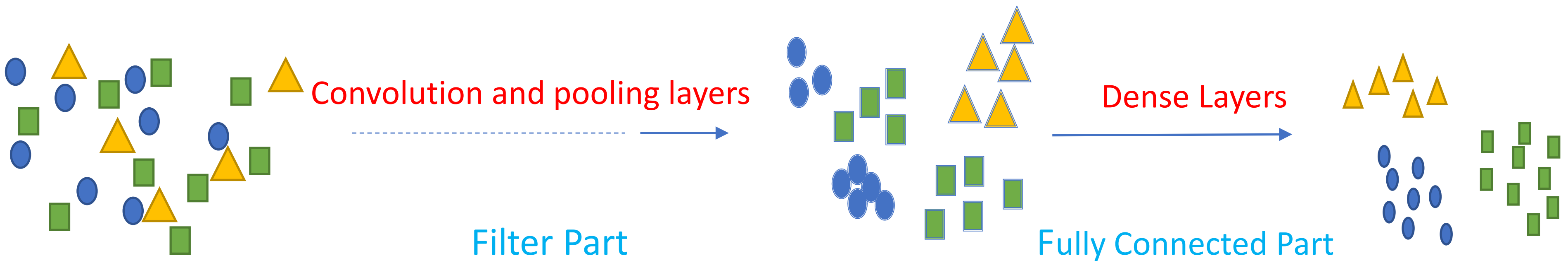
$$SI = 0(no\ sepration)$$

Extended SI: The “*k*th order of SI (k=1,2..)” denotes the relative number of all possible examples whose “*k*” nearest neighbors (in Lp norm) have the same labels with them.

What does it mean when we say the SI is perfect?

- When the separation index (for training examples of a classification problem) is perfect and the diversity of the training examples is sufficiently high, the classification problem can be solved by a high generalization.
- It means that any new example with a certain class is expected to be nearer to examples of the same class than to examples of other classes.
- In deep learning NNS, the separation index increases layer by layer through using convolution and pooling layers.
- Actually, in deep learning NNS, disturbances and distortions as well as the common features among examples of different classes are decayed layer by layer and concurrently, exclusive features of different classes dominantly make the vector space.

An illustrative examples



SI increase layer by layer

Applications of using Separation Index.

1. To Evaluate Different Datasets.

2. To Evaluate the performance of different layers of any Deep Learning NN.

3. To compress and reduce the redundancies of deep learning neural networks both in Filter part and Fully connected layers.

4. To learn and design Deep learning NNs in a forward manner by maximizing SI.
5. Actually parameters of each layer should be learned in order that the separtion index is maximized layer by layer.

Smoothness Index for Regression problems

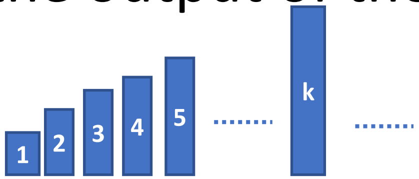
- Regression versus Classification
- Under Following conditions a classification problem can be interpreted as a regression problem:

(1) A variant of nearness among different classes is defined.

(2) the number of classes converges to infinity.

- About the condition (1), the examples of each class make a certain quantized level and actually near classes will have near quantized levels.

- About the condition (2), the number of quantized levels is increased to infinity and the output of the classification problem becomes purely analogue.



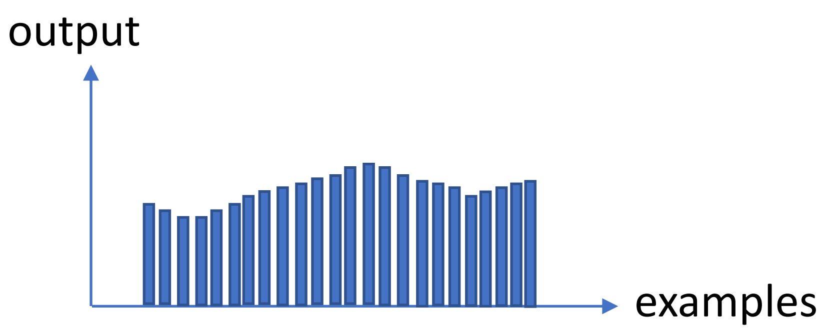
Some important notes

1. The quantized levels in a regression problem are the same labels we had in a classification problem.
2. Based on the dimension of the input, the output of a regression problem is like an analogue hyper-surface.
3. Like we stated (in a classification problem) that a perfect SI means that all examples of a class become sufficiently near together , here we state that examples of near classes become nearer together.
4. As a result, the corresponded near quantized levels become nearer together which lead to the concept of smoothness on the analogue output.
5. According to above-mentioned notes

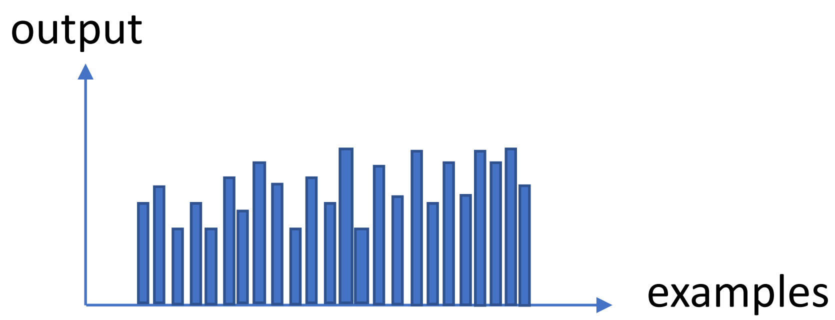
for regression problems we define smoothness index.

The concept of Smoothness index (SmI)

- We say that input-output data points $\{(\boldsymbol{x}^q, y^q)\}_{q=1}^{Q}$ in a regression problem make a high sufficiently smooth map if corresponded outputs of each set of near inputs (in Lp norm) become near together (in Lp norm).
- Two simple illustrative examples of smooth and non-smooth maps are shown as follows:



Smooth map



Non-Smooth map

How do we define the SmI in a Regression problem?

- Assume we have Q (n-dimensional) input-output data points $\{(\boldsymbol{x}^q, \boldsymbol{y}^q)\}_{q=1}^{Q}$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are two arbitral dimensional vectors.

1. Linear SmI is defined as follows:

$$SmI = \frac{1}{Q}\sum_{q=1}^{Q} Relu\left(1 - \frac{\left\|y^q - y^{q^*}\right\|_{L_p}}{d^q}\right)$$

$$q^* = \underset{i\in\{1,..,Q\},i\neq q}{\arg} \min\left\|x^q - x^i\right\|_{L_p}$$

$$d^q = \frac{1}{Q-1}\sum_{i=1,i\neq q}^{Q}\left\|y^q - y^i\right\|_{L_p}$$

$$0 \leq SmI \leq 1$$

2. Exponential SmI is defined as follows:

$$SmI = \frac{1}{Q}\sum_{q=1}^{Q} exp\left(-\frac{\left\|y^q - y^{q^*}\right\|_{L_p}}{d^q}\right)$$

$$0 \leq SmI \leq 1$$

What does it mean when we say the SmI is perfect?

- When the smoothness index (for training data points of a regression problem) is perfect and the diversity of the training data points is sufficiently high, the regression problem can be solved by a high generalization.
- It means that any new data point is an interpolated point of training data points and consequently it can be predicted with high accuracy.
- In deep learning NNS, the smoothness index increases layer by layer through using convolution and pooling layers.
- Actually, in deep learning NNS, disturbances and distortions as well as the bumps among training data points are decayed layer by layer and finally a smooth map will be appeared which is ideal for prediction purpose.

Applications of using Smoothness Index.

1. To Evaluate Different Datasets.
2. To Evaluate the performance of different layers of any Deep Learning NN.
3. To compress and reduce the redundancies of deep learning neural networks both in Filter part and Fully connected layers.
4. To learn and design Deep learning NNs in a forward manner by maximizing SmI.
5. Actually parameters of each layer should be learned in order that the smoothness index is maximized layer by layer.

Thank you