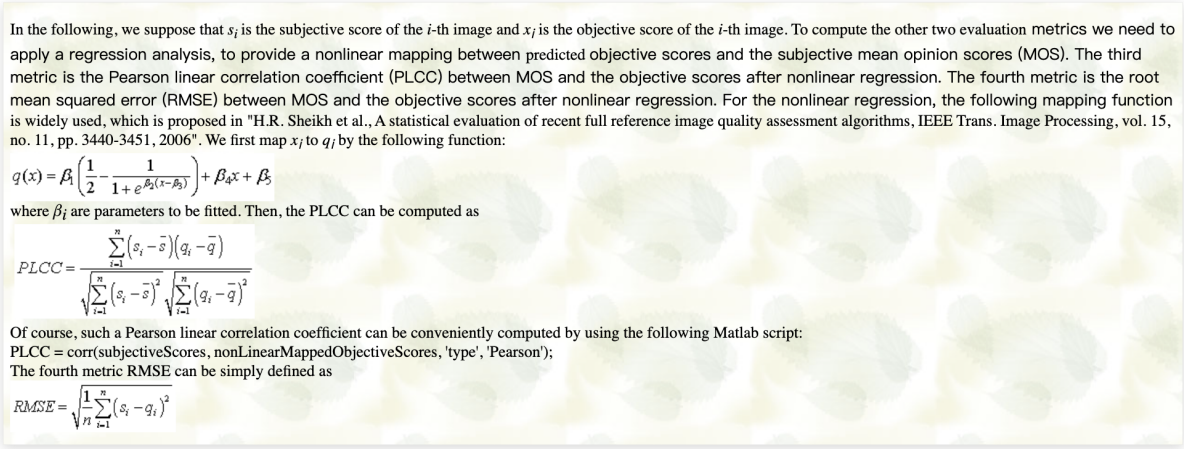
IQA Evaluation Metrics

(1) **Pearson Linear Correlation Coefficient (PLCC)**: we need to apply a regression analysis, to provide a nonlinear mapping between predicted objective scores and the subjective mean opinion scores (MOS).



In the following, we suppose that s_i is the subjective score of the i-th image and x_i is the objective score of the i-th image. To compute the other two evaluation metrics we need to apply a regression analysis, to provide a nonlinear mapping between predicted objective scores and the subjective mean opinion scores (MOS). The third metric is the Pearson linear correlation coefficient (PLCC) between MOS and the objective scores after nonlinear regression. The fourth metric is the root mean squared error (RMSE) between MOS and the objective scores after nonlinear regression. For the nonlinear regression, the following mapping function is widely used, which is proposed in "H.R. Sheikh et al., A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Processing, vol. 15, no. 11, pp. 3440-3451, 2006". We first map x_i to q_i by the following function:

$$q(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5$$

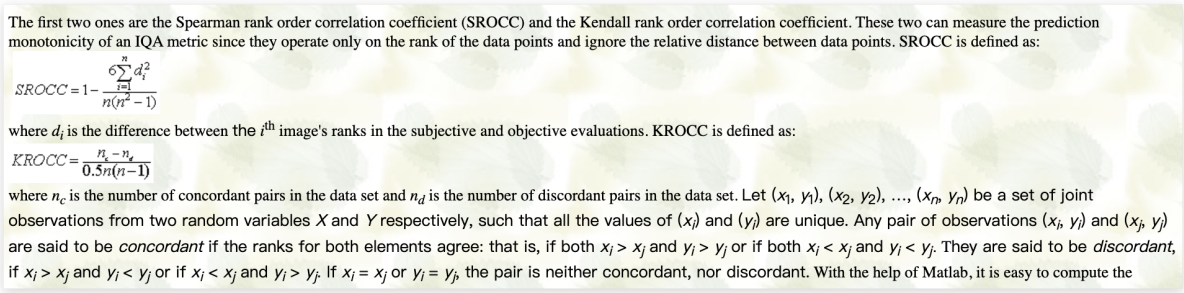
where β_i are parameters to be fitted. Then, the PLCC can be computed as

$$PLCC = \frac{\sum_{i=1}^{n} (s_i - \bar{s})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^{n} (s_i - \bar{s})^2} \sqrt{\sum_{i=1}^{n} (q_i - \bar{q})^2}}$$

Of course, such a Pearson linear correlation coefficient can be conveniently computed by using the following Matlab script:
PLCC = corr(subjectiveScores, nonLinearMappedObjectiveScores, 'type', 'Pearson');
The fourth metric RMSE can be simply defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (s_i - q_i)^2}$$

(2) **Spearman Rank-order Correlation Coefficient (SRCC) and the Kendall Rank-order Correlation Coefficient (KRCC):** These two can measure the prediction monotonicity of an IQA metric since they operate only on the rank of the data points and ignore the relative distance between data points.



The first two ones are the Spearman rank order correlation coefficient (SROCC) and the Kendall rank order correlation coefficient. These two can measure the prediction monotonicity of an IQA metric since they operate only on the rank of the data points and ignore the relative distance between data points. SROCC is defined as:

$$SROCC = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

where d_i is the difference between the i^{th} image's ranks in the subjective and objective evaluations. KROCC is defined as:

$$KROCC = \frac{n_c - n_d}{0.5n(n-1)}$$

where n_c is the number of concordant pairs in the data set and n_d is the number of discordant pairs in the data set. Let (x_1, y_1), (x_2, y_2), …, (x_n, y_n) be a set of joint observations from two random variables X and Y respectively, such that all the values of (x_i) and (y_i) are unique. Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be *concordant* if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. They are said to be *discordant*, if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant, nor discordant. With the help of Matlab, it is easy to compute the

1. where nc is the number of concordant pairs in the data set and nd is the number of discordant pairs in the data set.
2. Let (x1, y1), (x2, y2), …, (xn, yn) be a set of joint observations from two random variables X and Y respectively, such that all the values of (xi) and (yi) are unique.
3. **Any pair of observations (xi, yi) and (xj, yj) are said to be concordant if the ranks for both elements agree: that is, if both xi > xj and yi > yj or if both xi < xj and yi <**

yj. They are said to be discordant, if xi > xj and yi < yj or if xi < xj and yi > yj. If xi = xj or yi = yj, the pair is neither concordant, nor discordant.

Spearman ρ (rho) coefficient is a nonparametric measure of statistical dependence between the rankings of two variable assessing how well this relationship can be expressed using the monotonic function.

Kendall's τ (tau) measuring the ordinal association between two measured quantities.

(3) **Outlier Ratio (OR)**

prediction consistency

the percentage of the number of predictions outside the range of ±2 times of the standard deviations.

(4) **Scatter Distribution Plot of Subject MOS and Predictive scores**

In addition to the four evaluation metrics defined above, usually researchers plot the distributions of subjective scores and objective scores on a 2-D graph and also plot the fitted curve on the same figure. The following figure is such an example, which shows the scatter distributions of subjective MOS versus the predicted scores obtained by the FSIM IQA metric and the VIF IQA metric on TID2008 database.