Locally Weighted Regression

Alvaro Uceda
Aunir
1. Clean up calibration set

- There are 207 redundant samples in the calibration set.
- There are 54 spectra in validation set that have identical match in the calibration set.

2. Pre-process calibration and validation sets

- Savitsky-Golay derivative (1,15,2) and SNV was found to give the best results.
3. Calculate distances

- Un-scaled mahalanobis distances in the PCA score space (10 factors) of calibration spectra were used.

- Distance from each sample in validation set was calculated to the nearest samples in the score space and $nn=150$ (local) samples were found to be optimal and were indexed for each validation sample.
4. Remove extreme Y values

- A Global PLS was applied.
- Samples with extreme reference values with respect to the estimated values ($y_{est}$) were removed.
- A wide window ($> y_{est} - 30$ & $< y_{est} + 30$) was used for pruning.
5. Apply a weight function to local samples

- A cosine weighting function \( w \) was used instead of the traditional tricubic function in order to weight the local samples in proportion to their scaled distances \( d \):

\[
w = \cos \left( \frac{\pi}{2} d \right); \quad |d| < 1
\]

6. PLS regression on local samples

- A PLS regression (11 factors) was performed on weighted local samples to predict each validation sample.
Results

MAD = 0.32

Graph 1: A scatter plot showing the relationship between predicted and expected values.
Graph 2: A histogram showing the distribution of values.
Thank you

ALVARO UCEDA
alvaro.uceda@abagri.com