

Understanding Nonlinear Model Behavior with Shapley Values and Variable Sensitivity Measures

Sean Roginski¹, Barry M. Wise¹, Manuel A. Palacios¹
and Rasmus Bro²

¹Eigenvector Research, Inc.

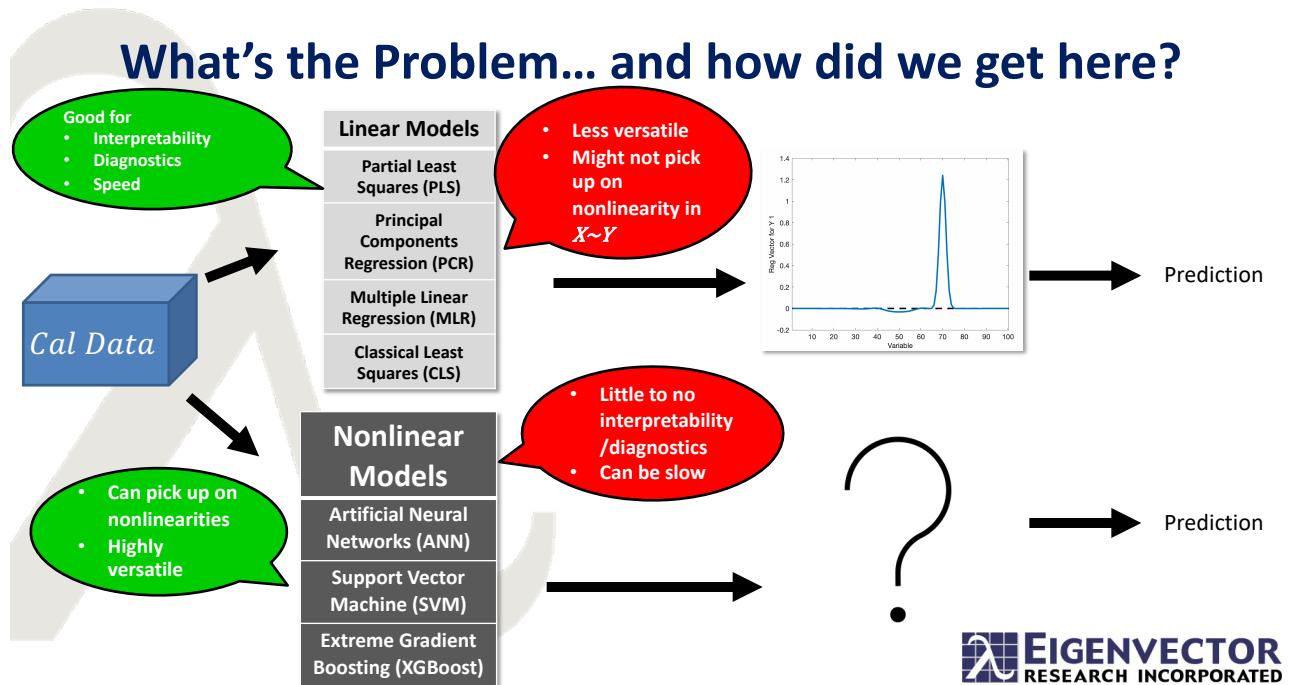
²University of Copenhagen



Outline

- What's the problem?
- What tools do we have to interpret regression models, and how do we do it?
- Can we interpret models like Support Vector Machines and Artificial Neural Networks?
- Tests and Discussion
- Conclusions





Interpreting Linear and Nonlinear Models

- Always important to know model behavior
- Important variables can be used in Variable Selection
- Avoid the 'blackbox paradox' with Explainable AI
- Interpret?

- Get a sense of how the model is generating the prediction.
- What variables is it using?
- Is it using variables we expect it to use?
- How robust is it to new artifacts or perturbations?

Variable Importance Measures

- Regression Vector
- Variable Importance in Projection (VIP)
- Selectivity Ratio

These **analyze properties** of the model

These are **subject to model type**

Sensitivity Tests

- Interference Test
- Single Variable Test
- Shapley Values

These measure **sensitivity of variables** in the model

Can be used for **any model type**

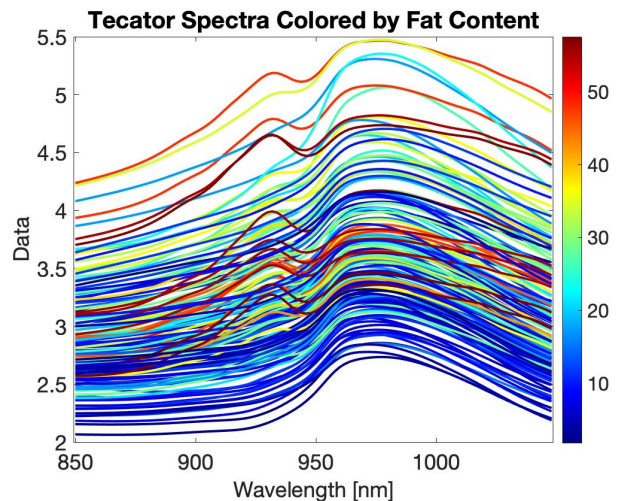


Tecator NIR Calibration Example

240 finely chopped pure meat samples measured by Infratec Food Analyzer. 100 wavelengths over 850-1050 nm.

Nonlinearity is exhibited between the spectra and the fat content.

ANN used to model this system for this example

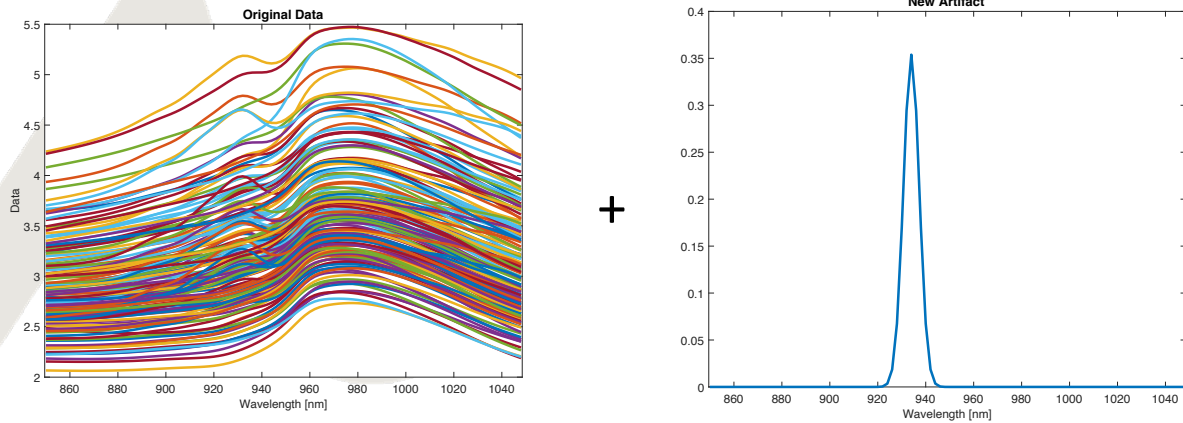


Borggaard, Thodberg, Analytical Chemistry, 64 (1992) 545-551.
Thodberg, IEEE Transactions on Neural Networks 7 (1996) 56-72.
<http://lib.stat.cmu.edu/datasets/tecator>

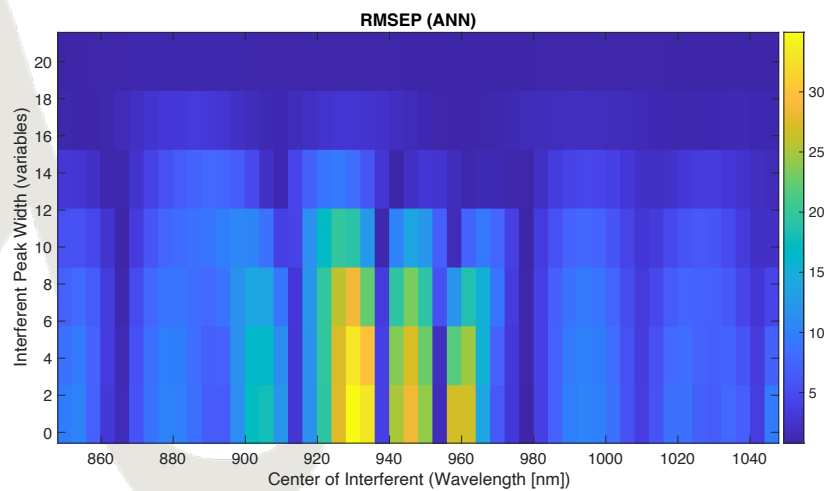


Interference Test

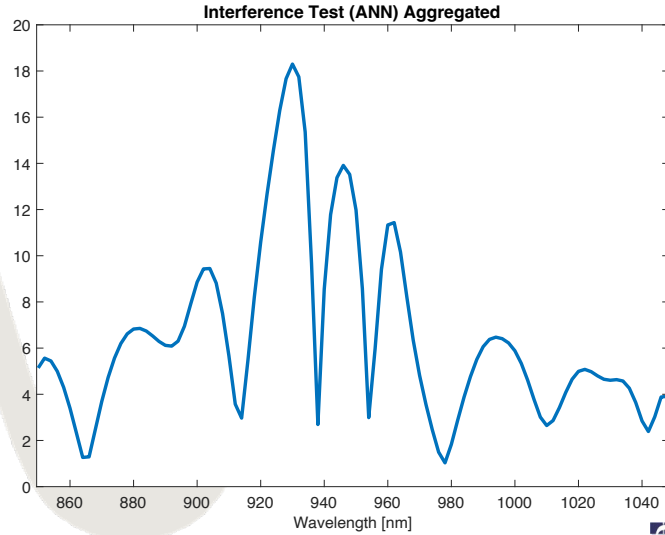
- Measures sensitivity of model to artifacts in new measurements
- Adds a Gaussian peak of varying width and position to X .
- If you have a chemical entity present in the test data not reflected in the calibration, what is the impact on predictions from the model?



Interference Test



Overall Response to Interferents

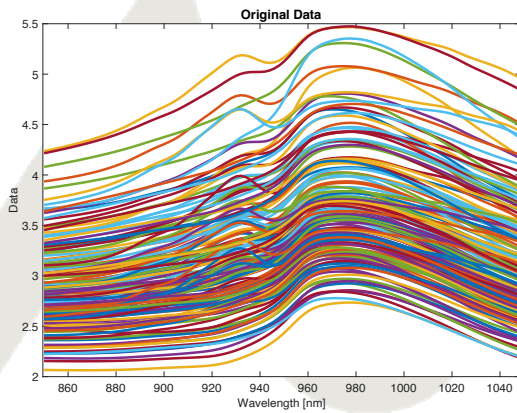


RMSEPs for average window width

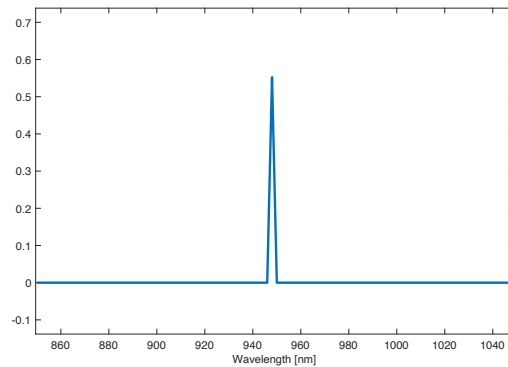


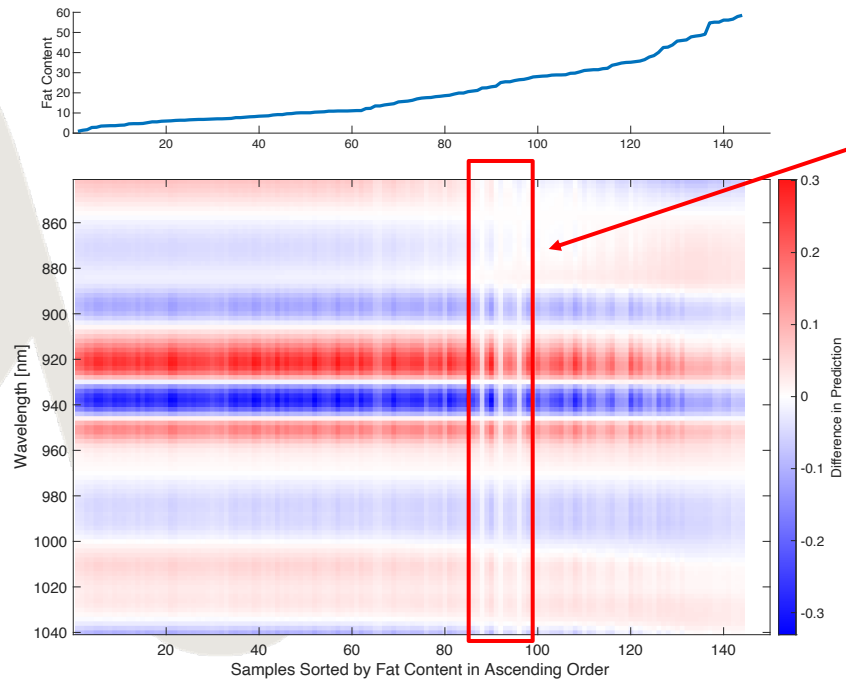
Single Variable Test

- Adds $\frac{std(j)}{100}$ for each variable j to X



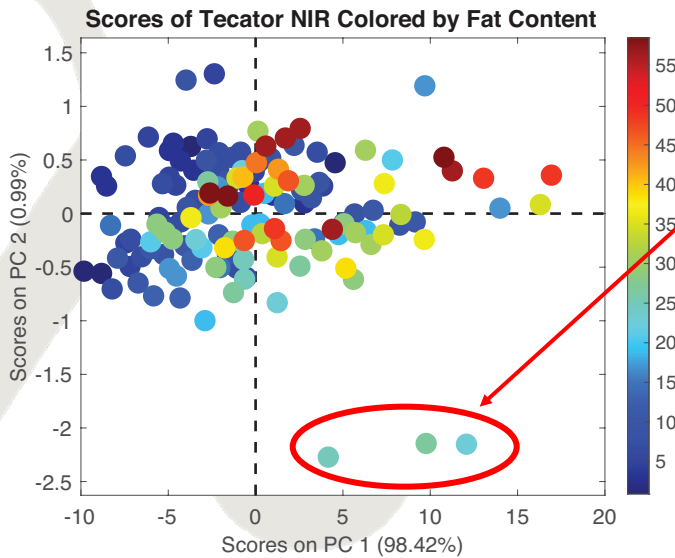
$std(948 \text{ nm})/100$





What's up with these three samples?

PCA Scores of Tecator

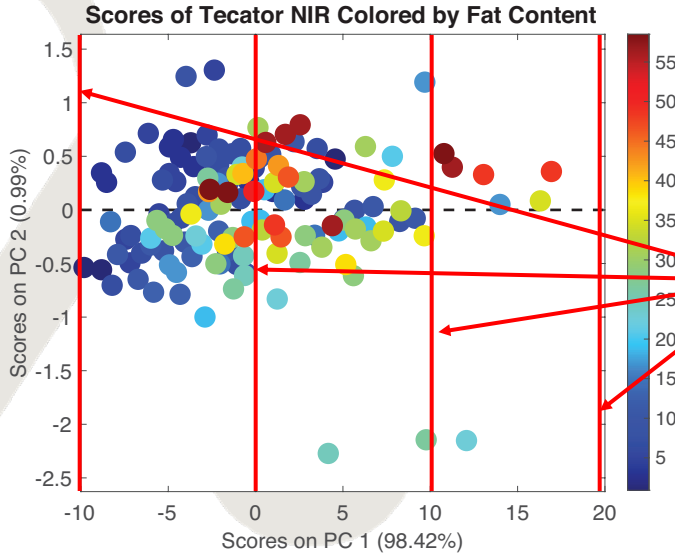


Model is insensitive around these three points

Suggests model is not generalizing here

“Learned the data?”

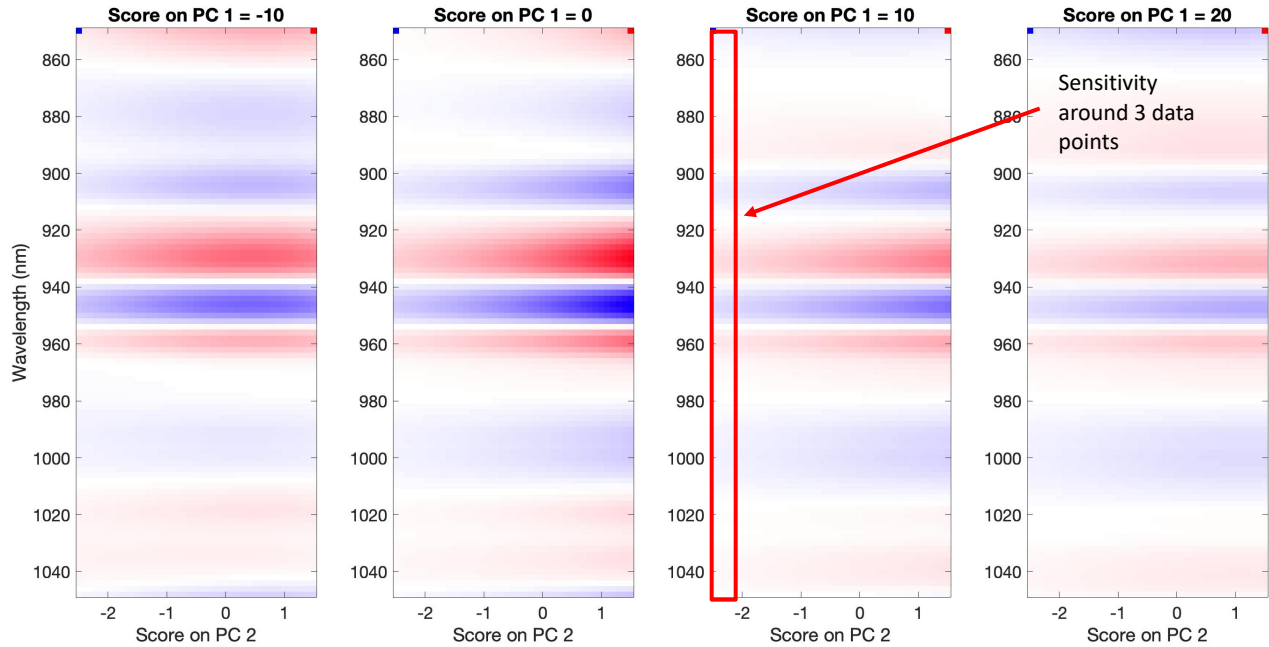
Sensitivity over PC1-PC2 Plane



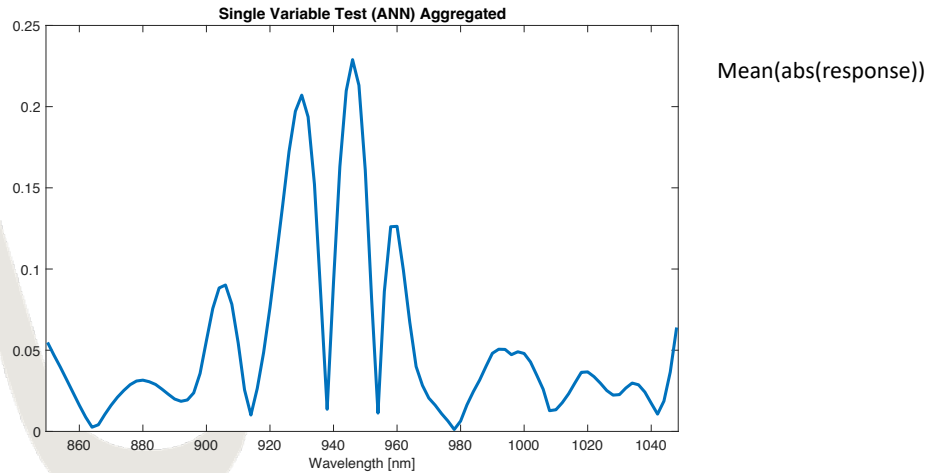
Generate synthetic data on grid over PC2 vs. PC1

Other PCs == 0

Show slices across PC1 at these locations



Overall Response to Sensitivity



Shapley Values

- From cooperative game theory
- Standard in Data Science
- Perturbations of the Xblock
 - Randomly sampling from random variables in the calibration dataset
 - Apply model to two perturbed samples at a time, take the difference
- Shapley Values assign individual effect of inclusion of a variable on a prediction, in units of the target
- Shapley Values are approximate



Lloyd Shapley, 2012 Nobel prize winner in Economics



Shapley vs. Sensitivity

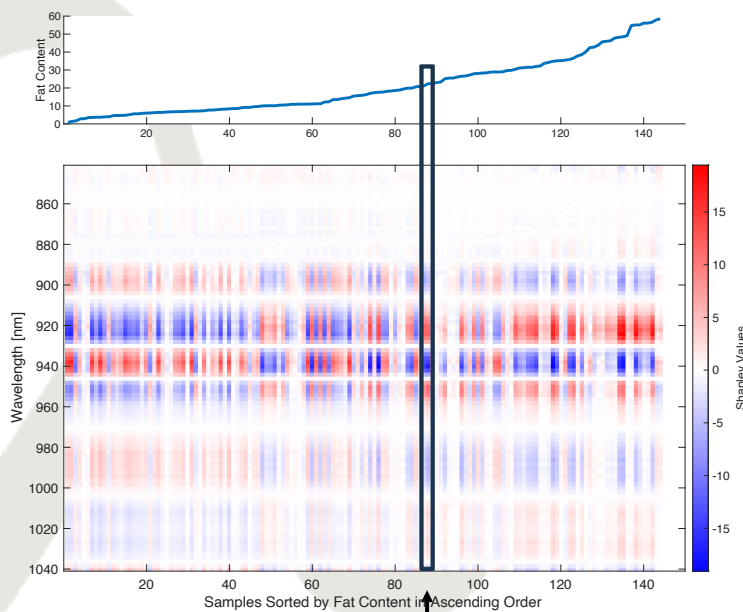
Shapley values and sensitivity tests answer different questions

The Shapley question:

For a given sample, what is the effect on the prediction of the *inclusion* of this variable?

The sensitivity question:

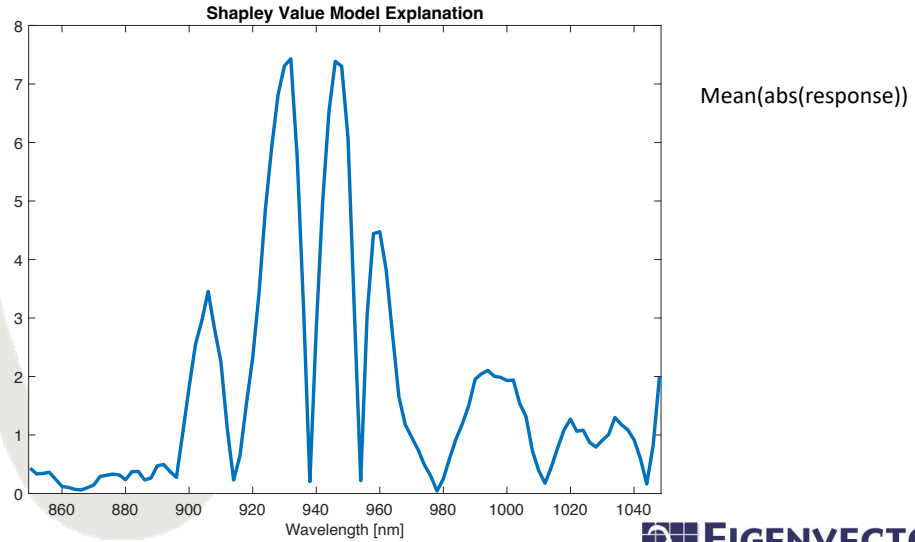
In the vicinity of a given sample, what is the effect on the prediction of a *change in stimulus* to this variable?



- Noted sample 87.
- The sum in the vertical direction (Shapley Values for this sample) plus the model intercept (mean y-value of 20% fat content) is equal to the sample's prediction (21.7%)
- The sum of the marginal contributions has to equal the total!



Average Shapley Value



Which is Best?

Regression Vectors
 • Subject to linear models

Interference Test
 • Model-Agnostic

VIP
 • Subject to PLS models

Single Variable Test
 • Model-Agnostic

Selectivity Ratio
 • Subject to PLS, MLR models

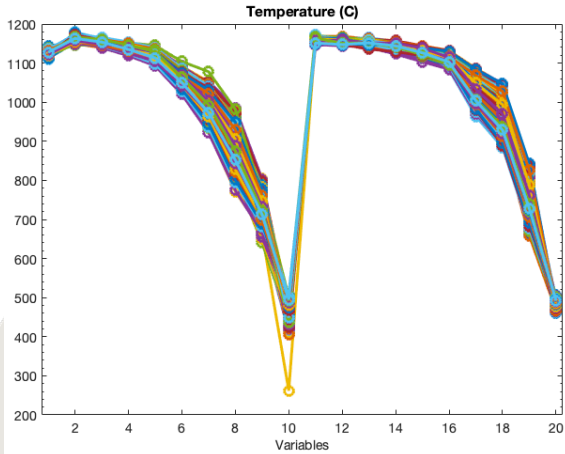
Shapley Values
 • Model-Agnostic

Which gives better interpretability?

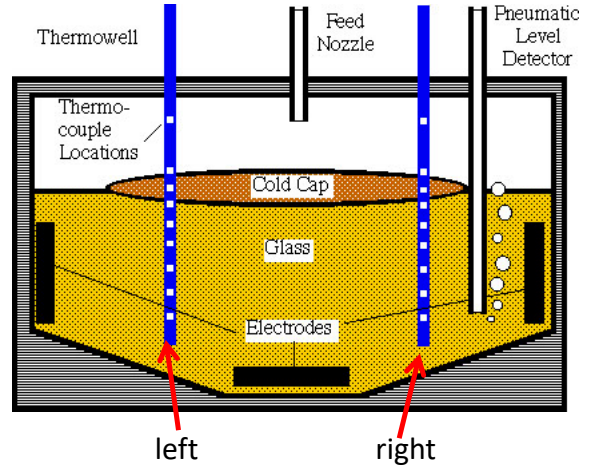
Which one gives best Variable Selection results?



Slurry Fed Ceramic Melter: SFCM



500 Samples, split 60-40 between calibration and validation sets
 Mean Center preprocessing

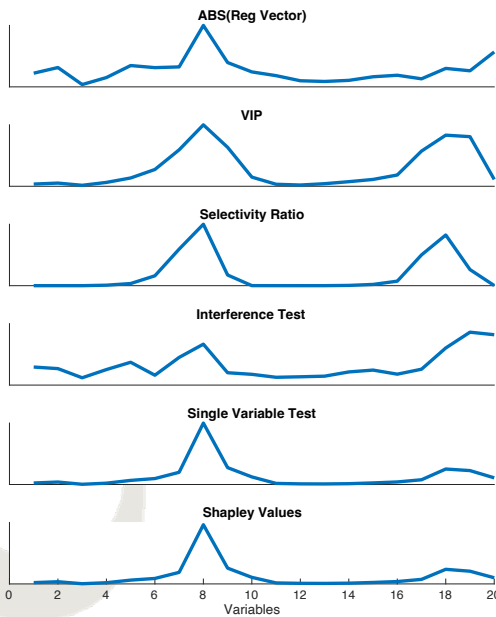


20 temperatures
 (numbered bottom to top)

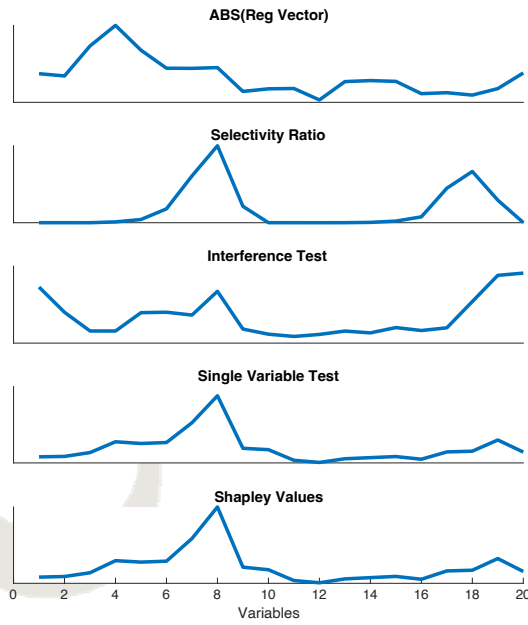


PLS

Model Explanation Results for PLS Model. RMSECV = 0.10727



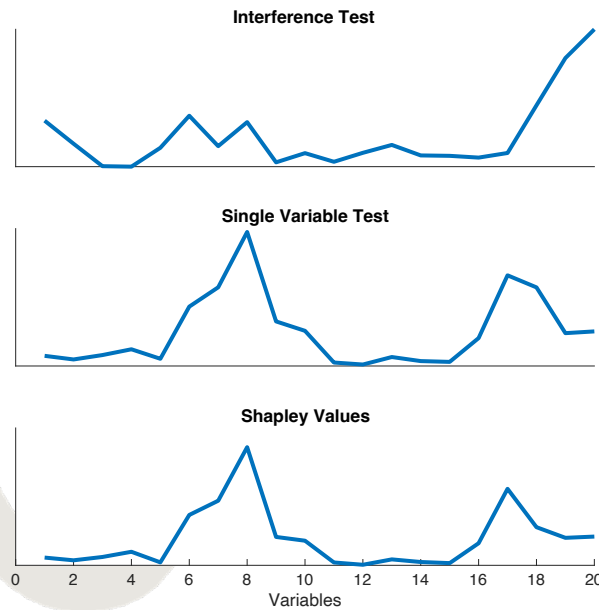
Model Explanation Results for MLR Model. RMSECV = 0.10975



MLR



Model Explanation Results for SVM Model. RMSECV = 0.10601

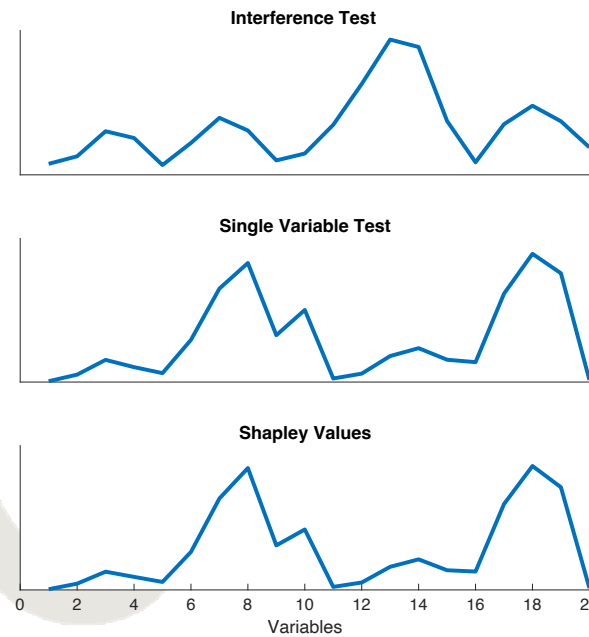


SVM



Model Explanation Results for ANN Model. RMSECV = 0.10883

25



ANN



26

Melter Data Results

- High overlap for all tools when interpreting each model type with the exception of the interference test
 - Most deemed Variables 7, 8, 9, 17, 18, and 19 to be important with slight differences
- Variable Selection helped with PLS, ANN, and SVM
- Single Variable Test and Shapley Values exhibit high overlap in the chosen variables and vector shapes
 - They also correspond closely with regression vector in PLS



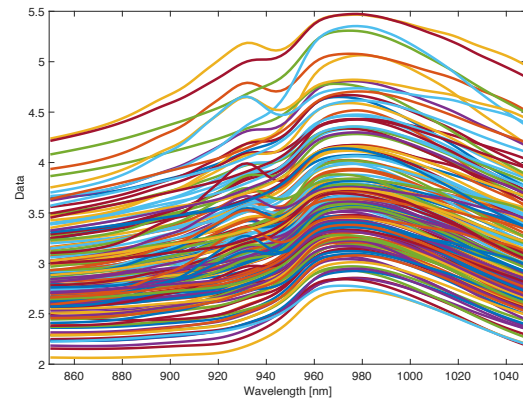
Tecator NIR Calibration Example

240 finely chopped pure meat samples measured by Infracore Food Analyzer. 100 wavelengths over 850-1050 nm.

Nonlinearity is exhibited between the spectra and the fat content.

1st Derivative, SNV, Mean Center preprocessing

'tecator' demo dataset is a 60%-40% calibration (144 samples)-validation (96 samples) split of the 240 samples in the published Tecator dataset

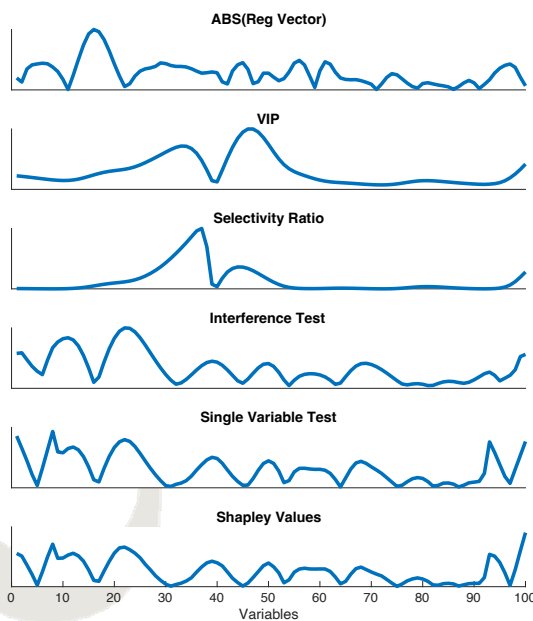


Borggaard, Thodberg, Analytical Chemistry, 64 (1992) 545-551.
Thodberg, IEEE Transactions on Neural Networks 7 (1996) 56-72.
<http://lib.stat.cmu.edu/datasets/tecator>

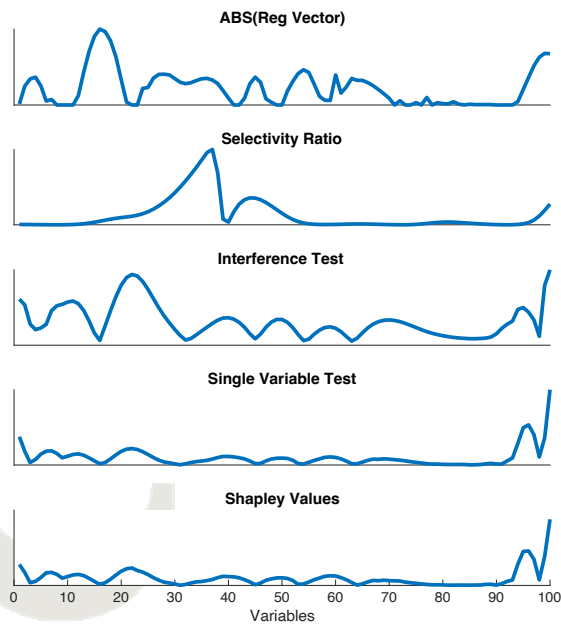


PLS

Model Explanation Results for PLS Model. RMSECV = 1.8347



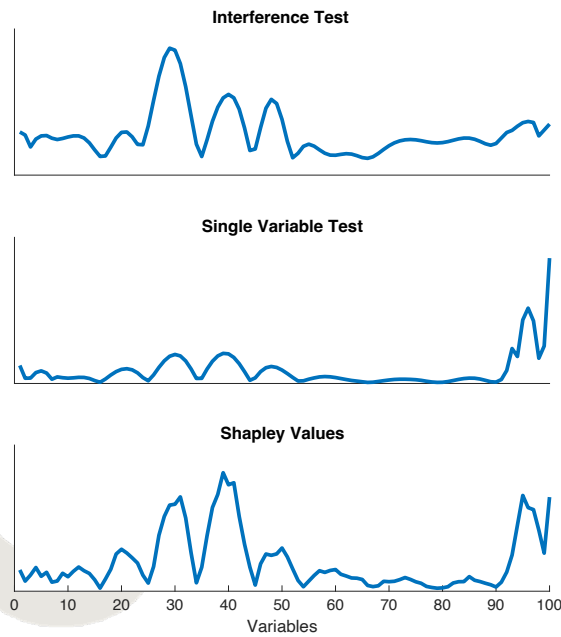
Model Explanation Results for MLR Model. RMSECV = 1.7936



MLR



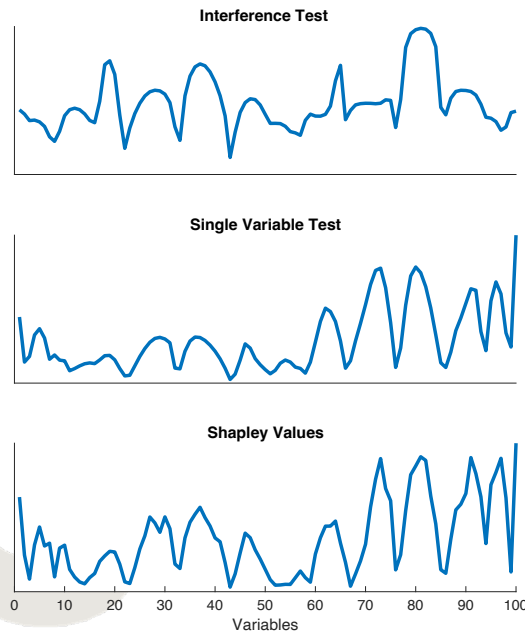
Model Explanation Results for SVM Model. RMSECV = 0.79059



SVM



Model Explanation Results for ANN Model. RMSECV = 0.56186



ANN



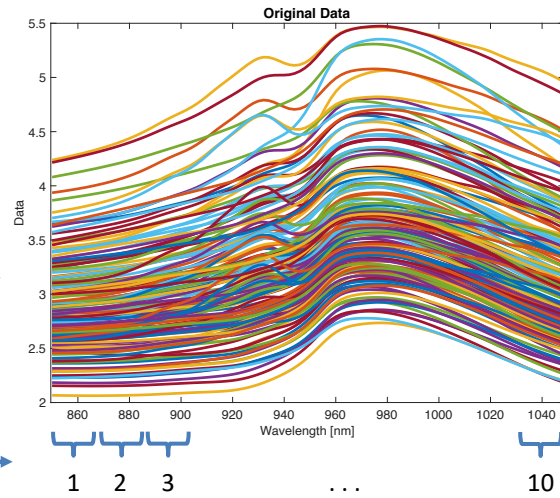
Tecator Results

- VIP and Selectivity Ratio give vastly different results compared to other tools. Doing variable selection with these gives poor results
- High overlap with all Sensitivity Tests, and they differ significantly with regression vectors.
- Variable Selection with these tools only improved in PLS, but also produced nice models with slightly higher RMSEPs that should be considered
- Single Variable Test and Shapley Values exhibit high overlap, except for ANN model

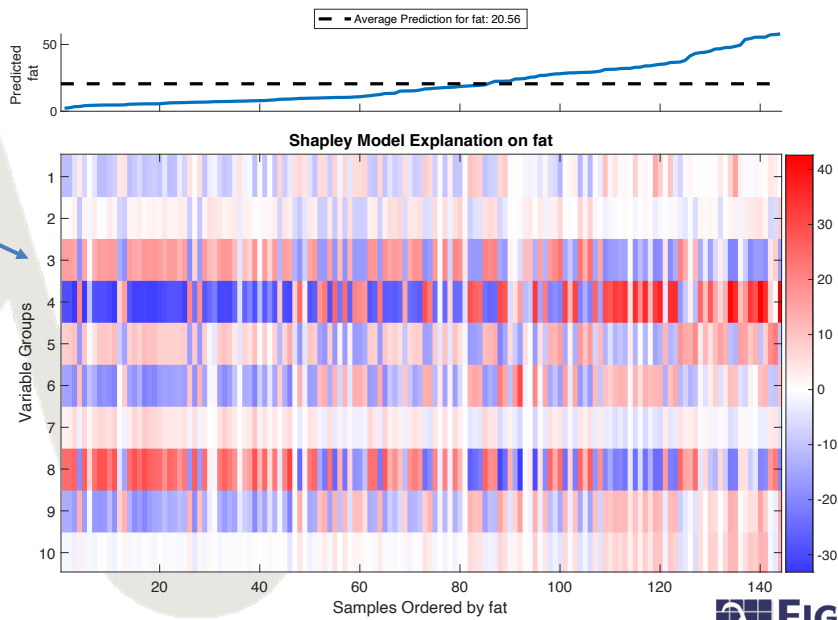


Shapley Values on Intervals

- Generation of synthetic samples assumes that variables in the data are independent, creating unrealistic data samples
- Variable groupings based on dependence is an alternative, then calculate Shapley Values for each group
- Significantly reduces computation time from 2^N down to 2^G for N variables and G groups of variables
- Example, Tecator with 10 groups of 10 variables instead of all 100 variables



Variables grouped into intervals of 10



Conclusions

- Variable Importance Measures can help interpret *linear* models
- Shapley and Sensitivity Tests can help explain *all* model types
- Overlap between Variable Importance Measures, Sensitivity Tests and Shapley
- All tools should be considered to interpret models
- Can also be used for Variable Selection to improve results