Evaluating Statistical Methods for Nuclear Forensics Analysis

Preliminary Examination

Arrielle Opotowsky

University of Wisconsin-Madison

29 January 2018

Outline

Nuclear Security and Forensics

background here of what it is

Needs in Nuclear Forensics

US/DHS stuff plus NF-specific needs

Contribution of Computational Methods

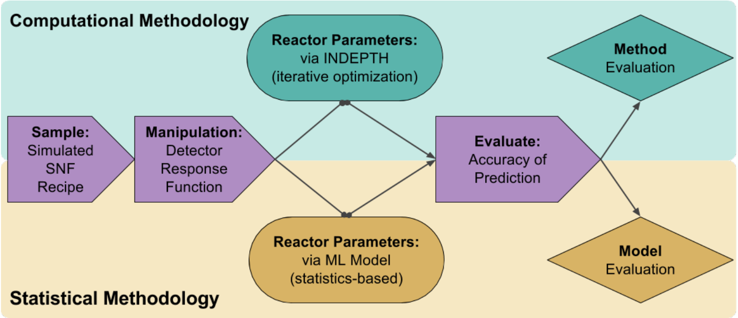


Figure 1: Nuclear forensics research: physical, experimental, and computational

Computational Methods

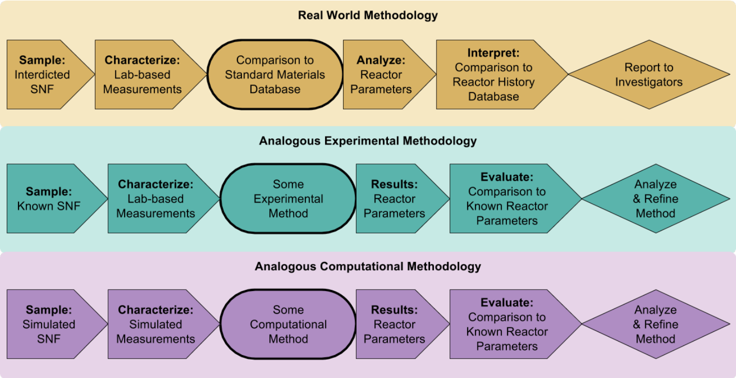


Figure 2: Comparison of computational approaches to nuclear forensics research

Contribution of Statistical Methods

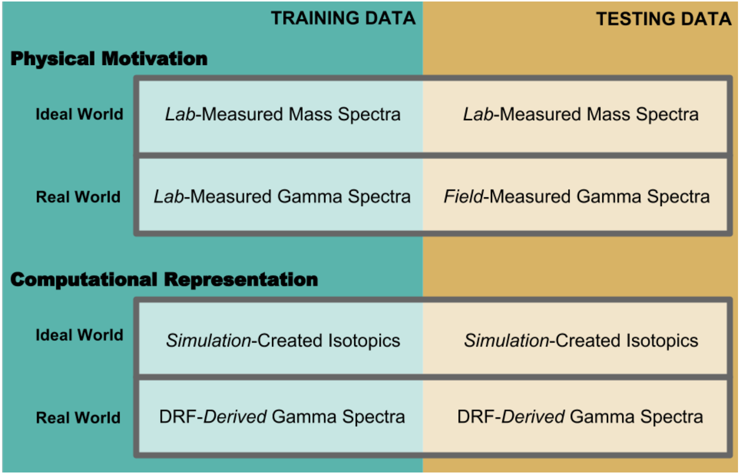


Figure 3: The benefits of data set modularity are easily implemented in this framework.

Statistical Methods

hi

Goal/Big Question

How does the ability to determine forensic-relevant spent nuclear fuel attributes using machine learning techniques degrade as less information is available?.

Outline

Technical Nuclear Forensics

something showing illicit trafficking. UOX seems to be the most common fear
SNF is really just for RDDs

Types of Investigations

pre-det v post-det

Nuclear Forensics as an Inverse Problem

inverse problem intro

Machine Learning

machine vs. statistical (domain knowledge-¿none)
supervised and unsupervised
clustering, dimensionality reduction
classification, regression – discrete and continuous variables

Vocabulary

labels
features
generalizability
prediction error
objective error metric v prediction error metric

Supervised Regression

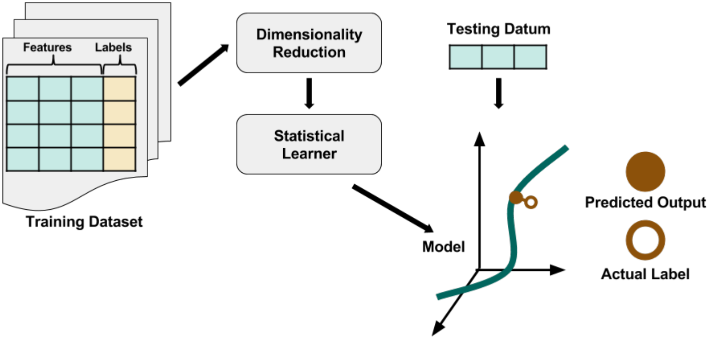


Figure 4: Schematic of a representative prediction workflow

Linear Models

Objective: minimize error over all training data (squared, absolute, etc)

$$F(\boldsymbol{X}) = \beta_0 + \sum_{j=1}^{p} x_j \beta_j \tag{1}$$

Regularization

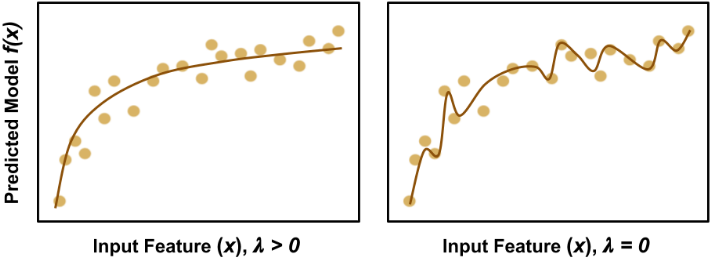


Figure 5: Schematic of the impact of regularization on the generalizability of an ML model

Nearest Neighbor Methods

Objective: minimum distance between test sample and training instance(s)

$$Y(\boldsymbol{X}) = \frac{1}{k} \sum_{x_i \in N_k(\boldsymbol{X})} y_i \quad (3)$$

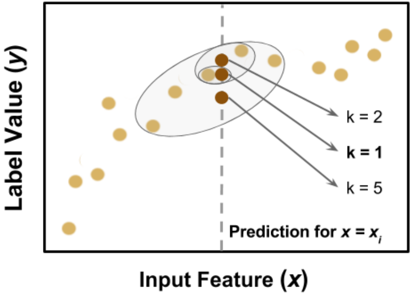


Figure 6: Illustration of the regularization effects by choosing k

Support Vector Machines

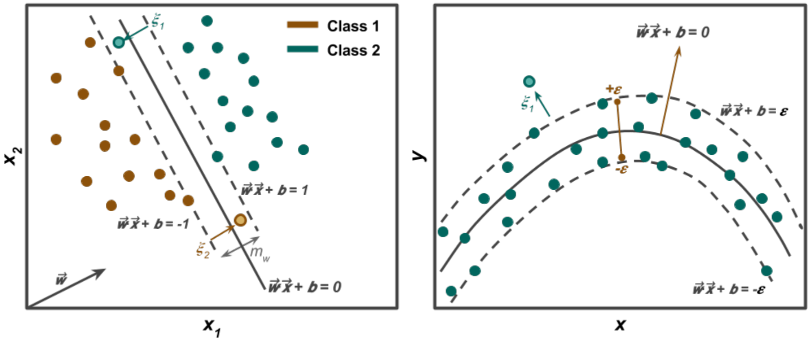


Figure 7: Classification with SVM and regression with SVR

Support Vector Regression with Many Dimensions

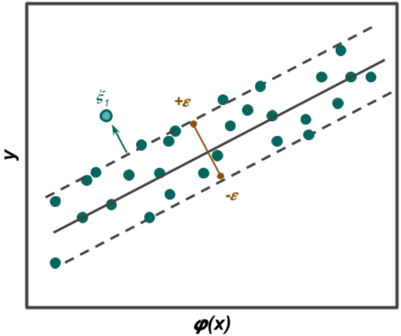


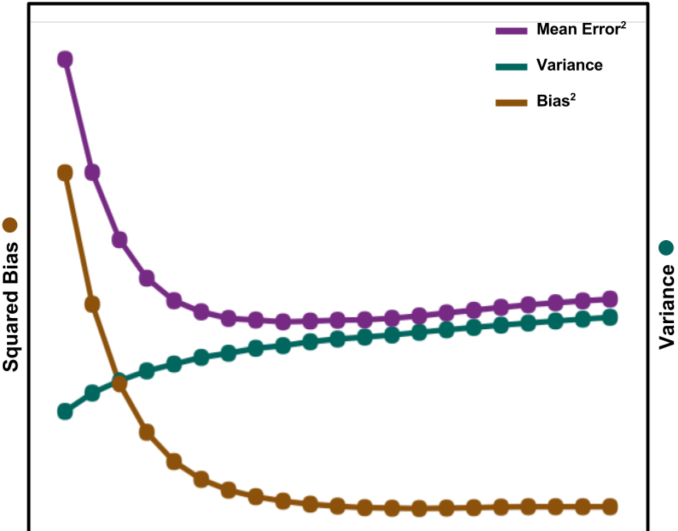
Figure 8: Diagram showing the use of the kernel trick with SVR

$$
\begin{aligned}
min\ \frac{1}{2}\|w\|^2 + C\sum_i \xi_i \\
subject\ to:\ |y_i - (w\phi(x_i) + b)| \leq \varepsilon + \xi_i \\
where:\ w = \sum_i \alpha_i y_i \phi(x_i) \\
and:\ K(x_i, x_j) = \phi(x_i)\phi(x_j) = e^{\gamma\|x_i - x_j\|^2}
\end{aligned}
\tag{4}
$$

Dimensionality Reduction

- Manual via domain knowledge or some measure
- PCA
- Factor Analysis
- ICA

Sources of Error



Types of Error

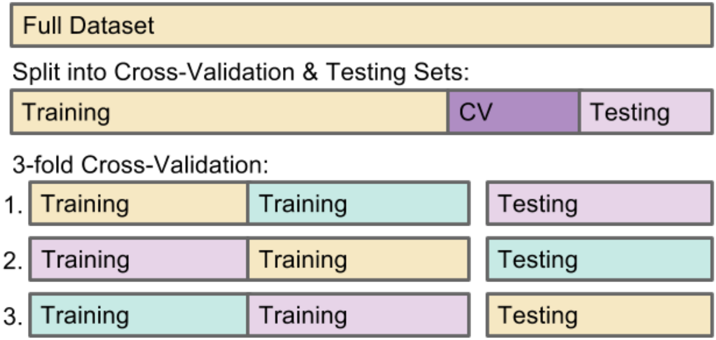


Figure 10: Diagram explaining the concept of k-fold cross-validation

Error Metrics

L1, L2, etc.
Used for model prediction error and obj funcs within the algorithms

Training Set Size: Learning Curves

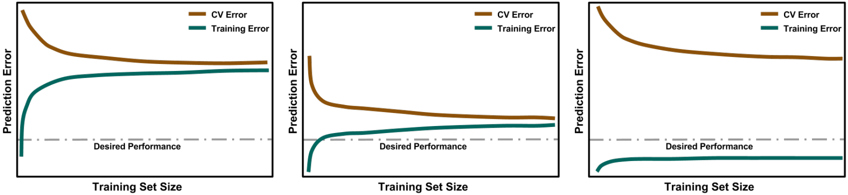


Figure 11: Learning curves for three training scenarios: high bias, balanced bias and variance, and high variance

Model Complexity: Validation Curves

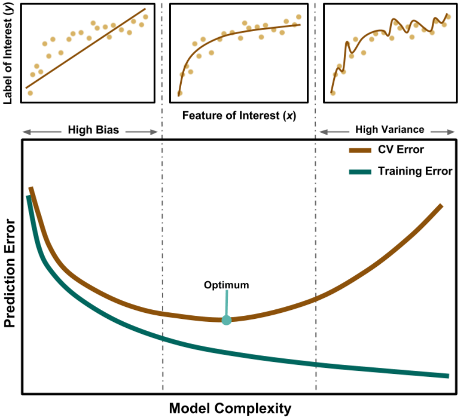


Figure 12: Validation curve showing different fitness of models

Model Comparison

Probabilities	Calculation Method and Example
P(D|M) Prior	*Given:* M = BWR U-oxide with burnup = ***x*** GWd/MTU *This is true:* D = nuclide vector with Pu-239 = ***y***% *With:* ***z***% probability *Calc'd from:* ORIGEN simulations in training set
P(M) Likelihood	*Given:* No direct information on D *This is true:* M = BWR U-oxide with burnup = ***x*** GWd/MTU *With:* ***z***% probability *Calc'd from:* Machine-learned model prediction
P(D) Marginal L.	*Given:* No direct information on M *This is true:* D = nuclide vector with Pu-239 = ***y***% *With:* ***z***% probability *Calc'd from:* Summation of training set instances
P(M|D) Posterior	*Given:* D = nuclide vector with Pu-239 = ***y***% *This is true:* M = BWR U-Ox with burnup = ***x*** GWd/MTU *With:* ***z***% probability *Calc'd from:* All quantities above

Table 1: Bayes

$$Posterior = \frac{Likelihood * Prior}{Marginal\ Likelihood} \tag{5}$$

Fuel Cycle Simulation

hi

Statistics Toolkit

hi

Computational Gamma Spectra

hi

Pre-detonation Materials of Interest

UOC
UOX powder
SNF
Reprocessed SNF

Statistical Methods Employed

Factor Analysis
SFCOMPO extension
Dayman paper on prediction ability wrt info reduction

Outline

Proposed Experiment Methodology

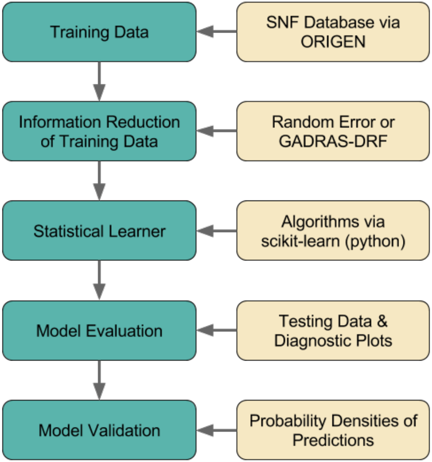


Figure 13: Workflow of the experiments with tools used for each step

SNF Simulations

hi

Information Reduction

hi

Algorithm Parameters

Algorithm	**Parameter**	**Value**
Nearest Neighbor Regression	*n*-neighbors	1
	Weights	uniform
	Distance Metric	L2: Euclidian Distance
Ridge Regression	Regularization, α	1.0
	Normalization	False
	Stopping Tolerance	0.001
Support Vector Regression	Kernel	Radial Basis Function
	Gamma, γ	0.001
	C	1000
	Epsilon, ε	0.1
	Stopping Tolerance	0.001

Table 2: caption

Initial Results

Algorithm	Error Origin	MAPE	RM [MW
Nearest Neighbor Regression	Testing Set	9.82	812
	5-fold Cross-Validation	2.24	421
Ridge Regression	Testing Set	15.68	104
	5-fold Cross-Validation	0.08	13.
Support Vector Regression	Testing Set	12.28	769
	5-fold Cross-Validation	2.08	188

ML Model Prediction with Reduced Information

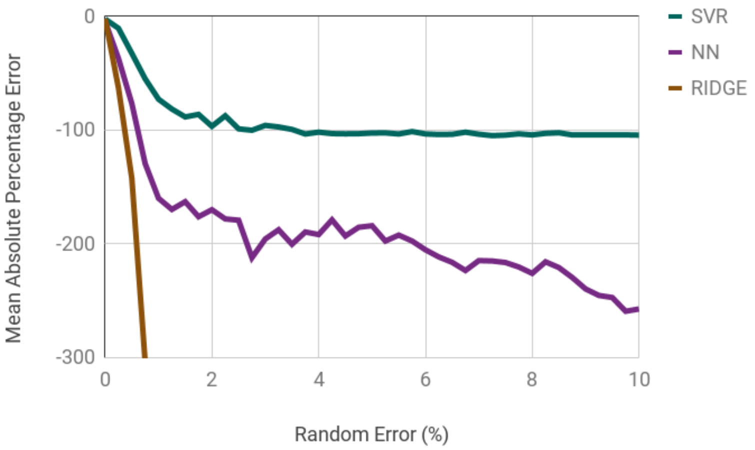


Figure 14: caption

SVR Learning Curve

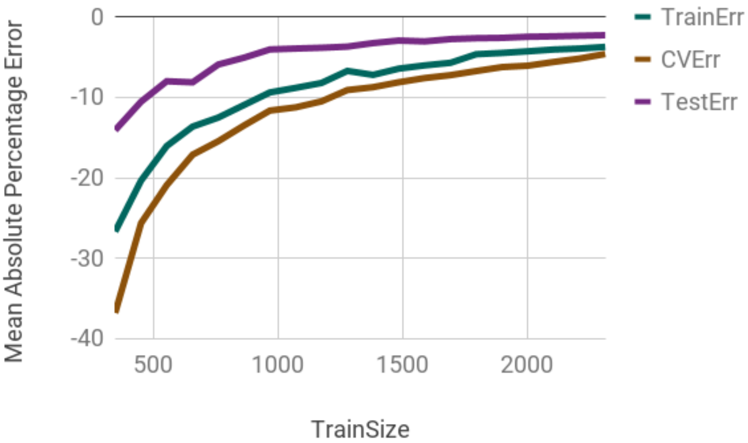


Table 4: caption

SVR Validation Curve

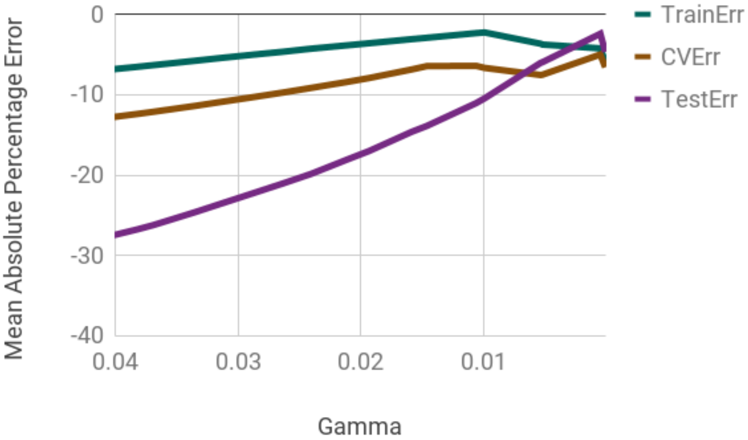


Table 5: caption

Outline

Research Proposal Preparations

Previous Work -¿ SFCOMPO-based
Finalizing set of algorithms
computational resources

Viability of Statistical Learning on Direct Isotopics

Purpose
Variables

Viability of Statistical Learning on Direct Isotopics

Hypotheses
Risks

Viability of Statistical Learning on Gamma Spectra

Purpose
Variables

Viability of Statistical Learning on Gamma Spectra

Hypotheses
Risks

Viability of Statistical Learning on Reprocessed Fuel

Purpose
Variables

Viability of Statistical Learning on Reprocessed Fuel

Hypotheses
Risks

Probability Distributions

Here, we change the meaning of the variables to represent probability distributions. C is a constant given by the marginal likelihood, which can be ignored when calculating relative probabilities, and $\boldsymbol{d}$ and $\boldsymbol{m}$ represent the training data set and model parameters, respectively. Thus, $P(\boldsymbol{d}|\boldsymbol{m})$ is the likelihood distribution function, $P(\boldsymbol{m})$ is the prior probability distribution, and $P(\boldsymbol{m}|\boldsymbol{d})$ is the posterior probability distribution.

$$P(\boldsymbol{m}|\boldsymbol{d}) = C \ * \ P(\boldsymbol{d}|\boldsymbol{m}) \ * \ P(\boldsymbol{m}) \tag{6}$$

Mathematically speaking, the distributions are obtained by integrating over the relevant probability densities. For example, the prior probability distribution can be calculated, where $\boldsymbol{m}$ is the range of predicted model parameters, i.e. burnup values, and $\boldsymbol{d}$ is a set of nuclide vectors. Also, here, $\rho(\boldsymbol{x}) = \prod_i \rho(x_i)$.

$$P(\boldsymbol{m}) = \int_{\boldsymbol{m}} \rho(\boldsymbol{d}) d\boldsymbol{d} \tag{7}$$

Similarly, the likelihood distribution function is obtained as in Equation

Estimating Density Functions

Convert wordage to graphic?

Posterior Odds

$$\frac{P(m_i|d)}{P(m_j|d)} = B_{ij}\frac{P(m_i)}{P(m_j)} \tag{9}$$

$\|\ln B_{ij}\|$	Probability	Likelihood Strength
< 1.0	< 0.750	Inconclusive
1.0	0.750	Weak
2.5	0.923	Moderate
5.0	0.993	Strong

Table 6: Model comparison using likelihood strength

Outline

References I