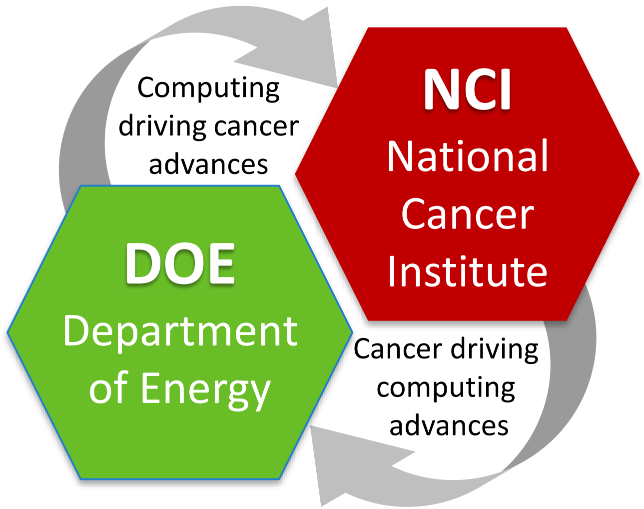
Crosscutting Technologies

Uncertainty Quantification



Tanmoy Bhattacharya

Oct 18, 2017

LA-UR-17-29458

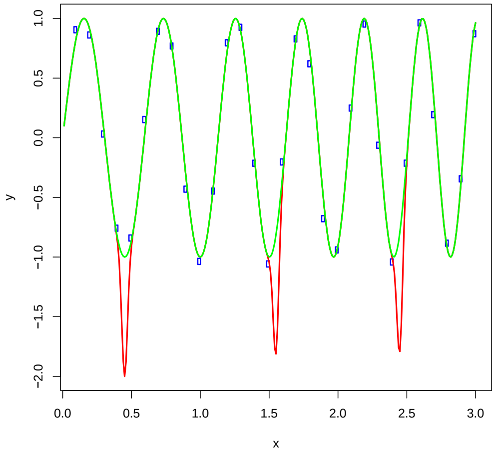
Why do we need uncertainty quantification?

- ***Machine learning* provides description of training data.**
  - Based only on input data with little expert knowledge.
  - Often opaque, based on subtle correlations.
  - Generalizes to similar data, but what is *similar* is not clear.
- **Data is currently**
  - Unimodal.
  - Collected opportunistically.
  - Has little gold-standard ground truth.
- **To reduce human workload,**
  - Need confidence in individual predictions: triage the highly certain cases.
  - Understand both statistical errors and data biases.
  - Quantify model transfer uncertainty.

UQ allows division of work between machines and humans

Generalization Error

- No amount of examples can predict an unseen point without assumptions.
- Function space is huge



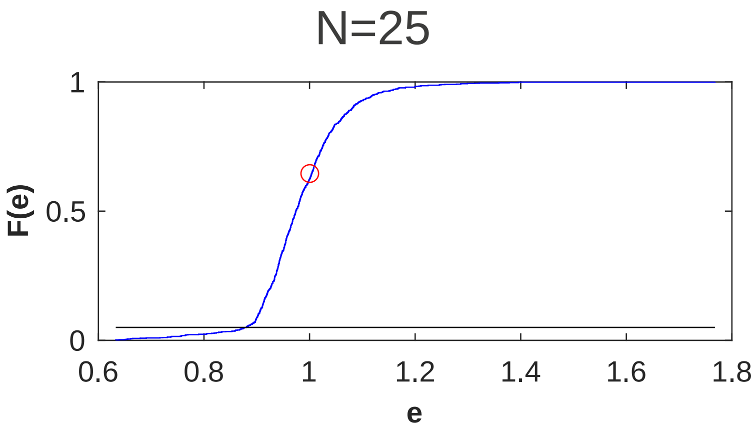
- Blue data: sin(10 x + x2) + random
- Green fit: sin(10x + x2)
- Red fit: sin(10x + x2) + repeated Gaussian
  - **Allow only if evidence very strong**
  - **Or if a repeated Gaussian is what we expect**

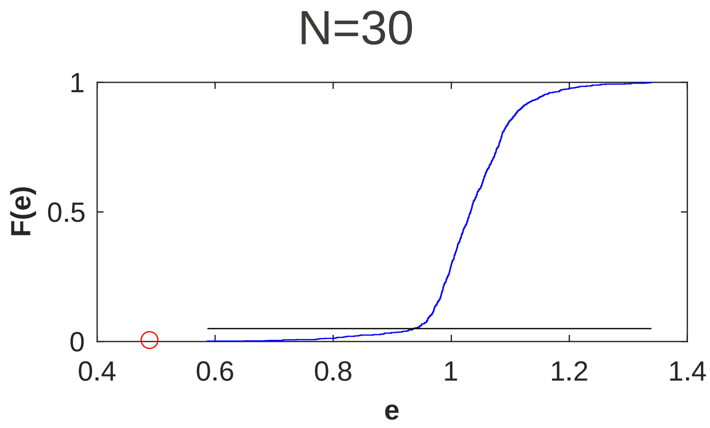
- Need to restrict to or favor parts of function space.
- Increase in data allows more complicated models without over-fitting.

There is an unavoidable tradeoff between ability to fit and prediction

Example: Checking if data has signal

- **Are we predicting *better than random?***
  - Even random data can be *predicted*, based only on frequencies
  - Remove all signal that one is interested in by permutation
  - Measure estimated error on this random data *using the same methodology*
  - Allows us to measure whether prediction is based on expected signal

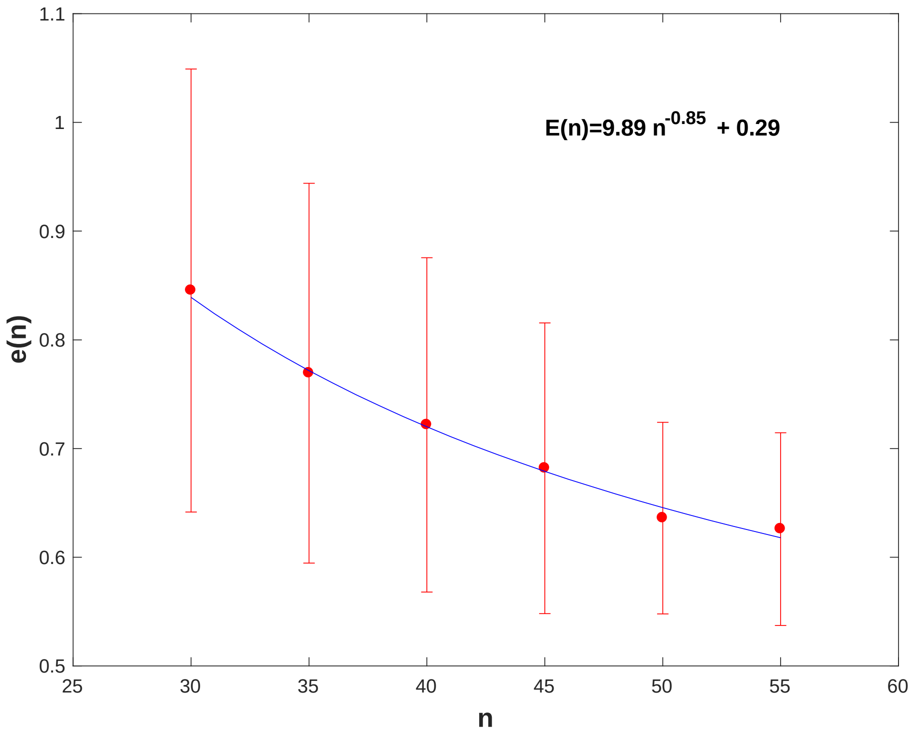




There is a prediction floor one reaches at low sample sizes

Example (continued): Empirical error curves

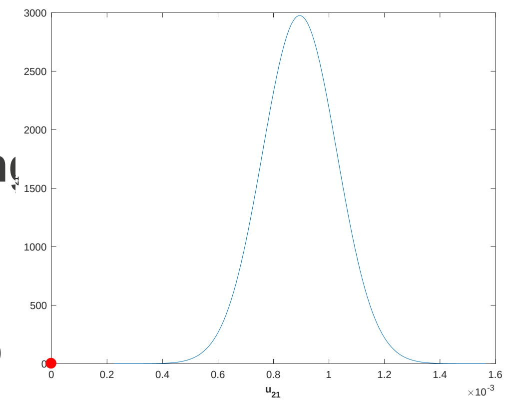
- **In simple machine learning techniques, error for large amount of data typically falls off as a power law.**
- **One can measure this error for different sample sizes.**
- **This curve can be extrapolated to estimate the *oracle error*: the amount of error that is intrinsic to the method.**

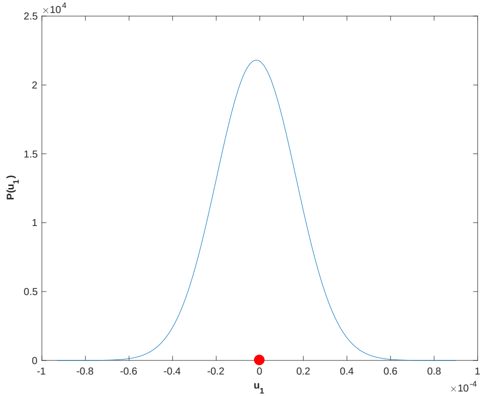


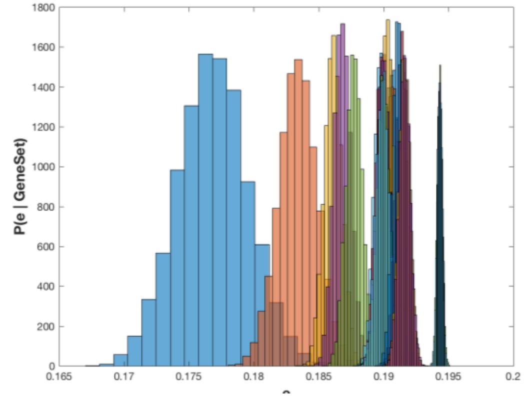
For simple machine learning, there is an error floor at large sample sizes

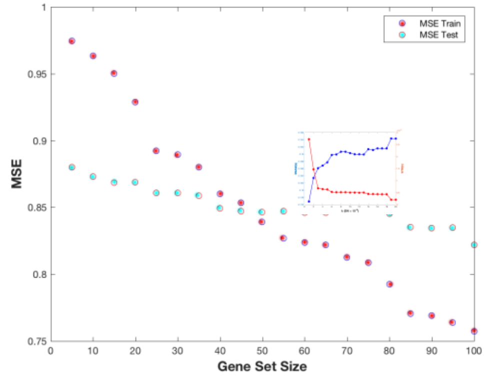
Example (continued): Model complexity

- **In standard machine learning, one can control model complexity by doing a *variable selection.***
- **As more variables included, fitting better, so *train set error* reduces.**
- ***Test set error* stabilizes**
- **As model complexity increases**
  - Each training set is fit better
  - Different training sets give different models





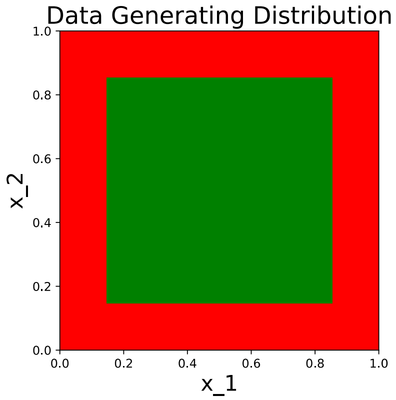


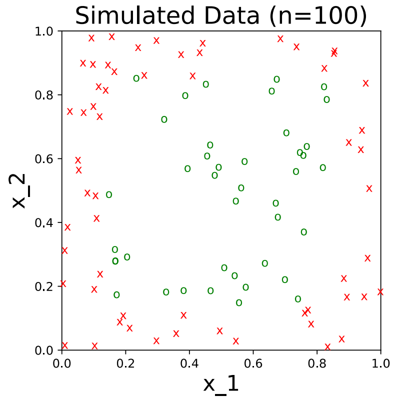


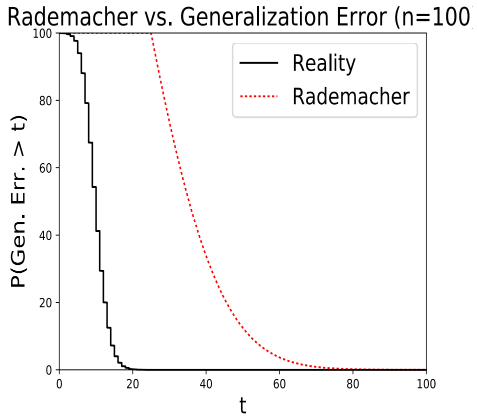
Bias variance tradeoff as model complexity increases.

Rademacher Error

- **Ideal situation**
  – Model predicts real data well
  – Model does not predict random data at all
- **Then, one can be sure that the prediction is *real*, and will generalize.**
- **Formalized in Rademacher bounds: strictly conservative upper bound.**
- **Bound becomes tight as data size increases**

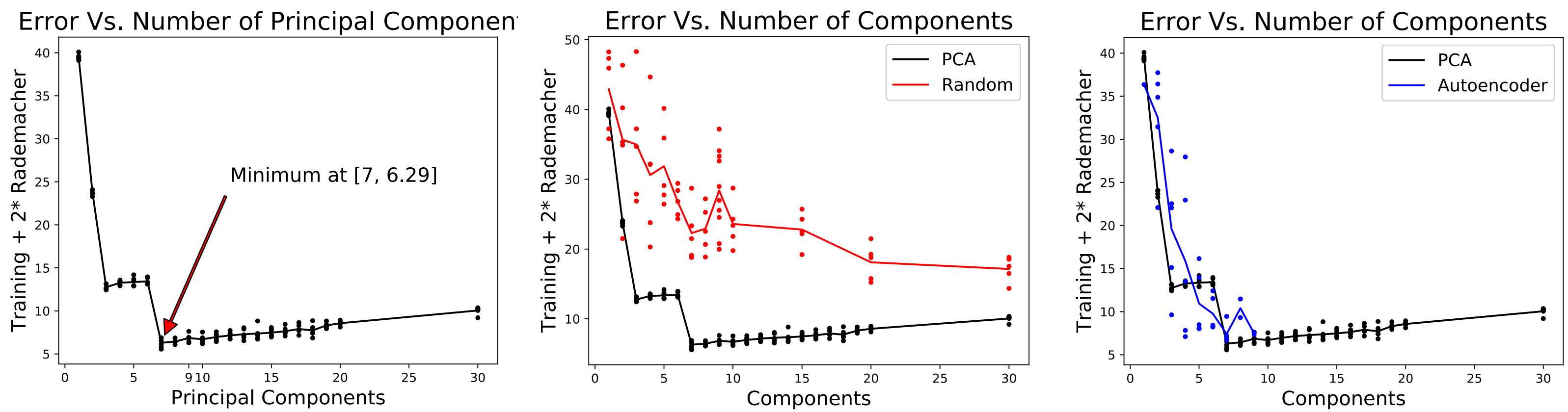






Example: Autoencoder or Principal Components

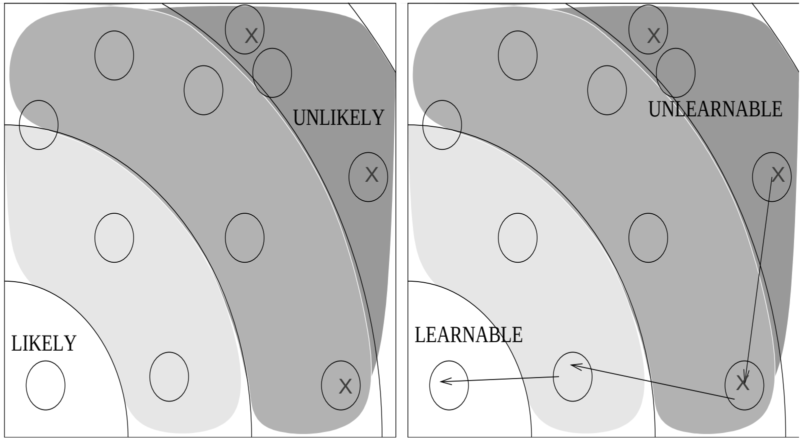
- **Compare**
  - Principal Components
  - Random Components
  - Autoencoder Components

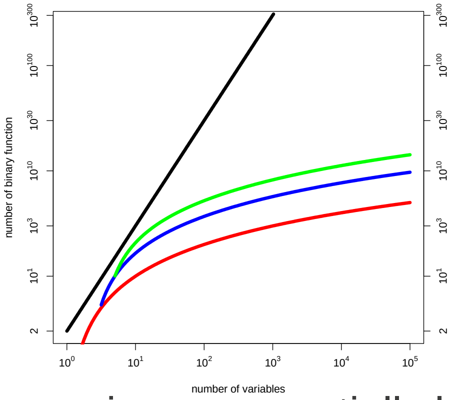


Principal Components and Autoencoders give similar Rademacher bound

Rademacher to bound deep learning?

- **Traditionally, one gets strong uncertainty guarantees using these**
- **Shown to *not work* for deep learning**
  – First memorize (really bad and unlearnable solution)
  – Optimize to find a better solution
- **What counts as better?**





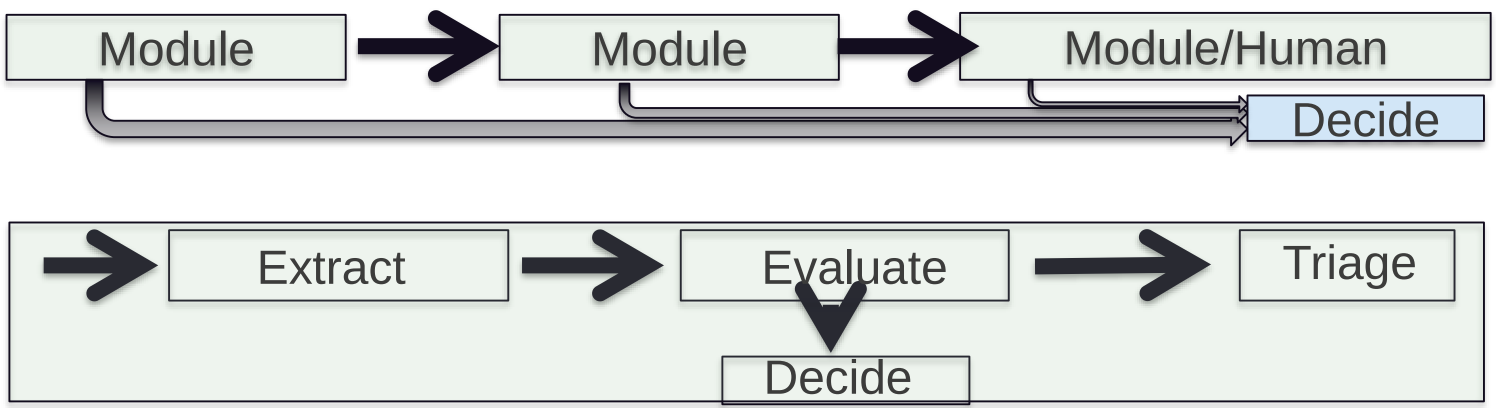
- Function space dimension exponentially large.
- Unreasonable effectiveness of learning IGUs.
- Theory allows *meta*learning across domains: model transfer uncertainty.
- UQ from correlations between generative and analytical models.

Work in progress to use other methods like dropout sensitivity

Uncertainty stratification and Triage

Pilot 3: Machine-assisted extraction

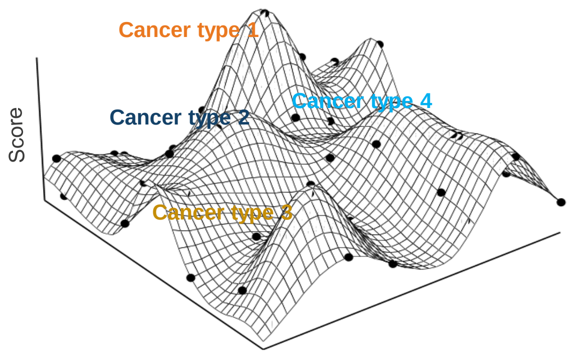
- **Measuring average uncertainty only a first step**
- **If we can separate certain and uncertain situations**
  - Can spend expensive resources on uncertain situations



UQ can be used to distill error-free output

UQ on individual instances

- **No assumption-free generalization guarantees**
- **Assume close by train cases inform uncertainty**
- **Assumptions dictate what is close by.**
  - Close by in input space: Similar word use, similar format, …
  - Close by in output space: Difficulty making a call, boundary of match region, …

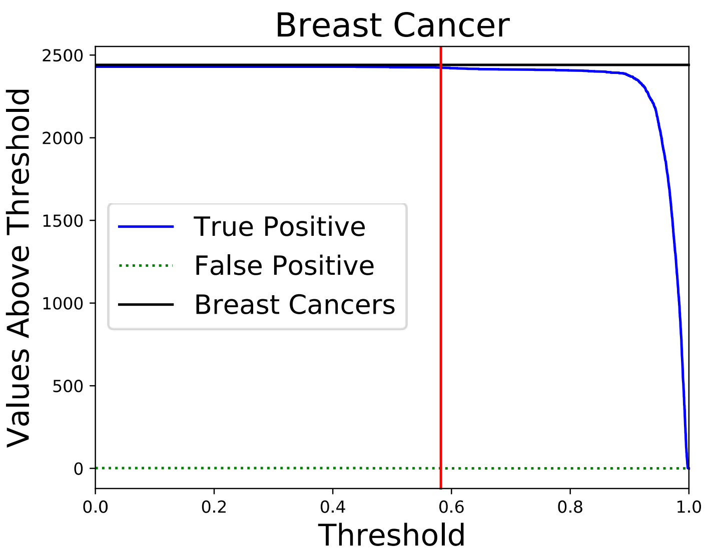


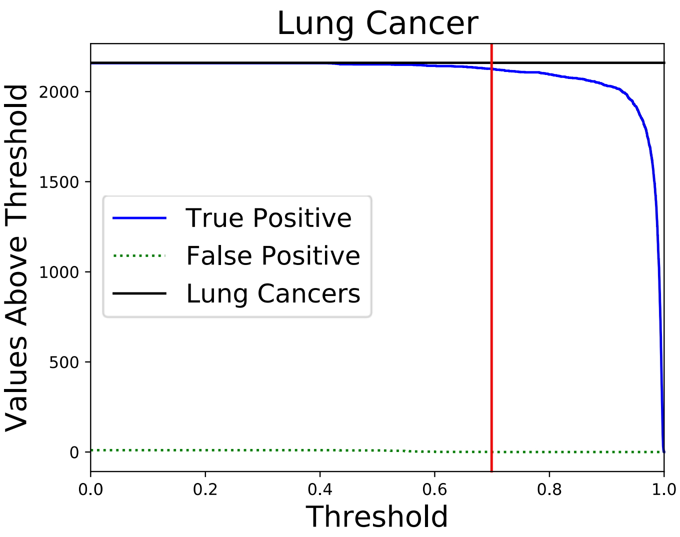
Report ID	Lung	Breast	Colon	Prostat e	…..
CT-REC-XXXX	0.68	0.12	0.09	0.02	…
HI-REC-XXXX	0.28	0.23	0.26	0.07	…

← Confident

← Not confident

Example: use highest score





Conclusions

- **Uncertainty quantification bounds errors on cases unseen**
  - Standard approaches available
  - Need modification for deep learning
- **Uncertainty quantification allows optimal design of experiments**
  - Simulations can address lacunae in knowledge
  - Effects of sampling biases can be quantified
- **Uncertainty quantification can allow certainty distillation**
  - Can provide a subset with negligible errors
  - Separate the easy cases from the hard cases

UQ methods in development here will help other deep-learning projects