

To access the tutorial documentation:

https://github.com/moovai/model_validation_tutorial



**Validate and monitor machine
learning models**

Olivier Blais

To access the tutorial documentation:

https://github.com/moovai/model_validation_tutorial

Objectives

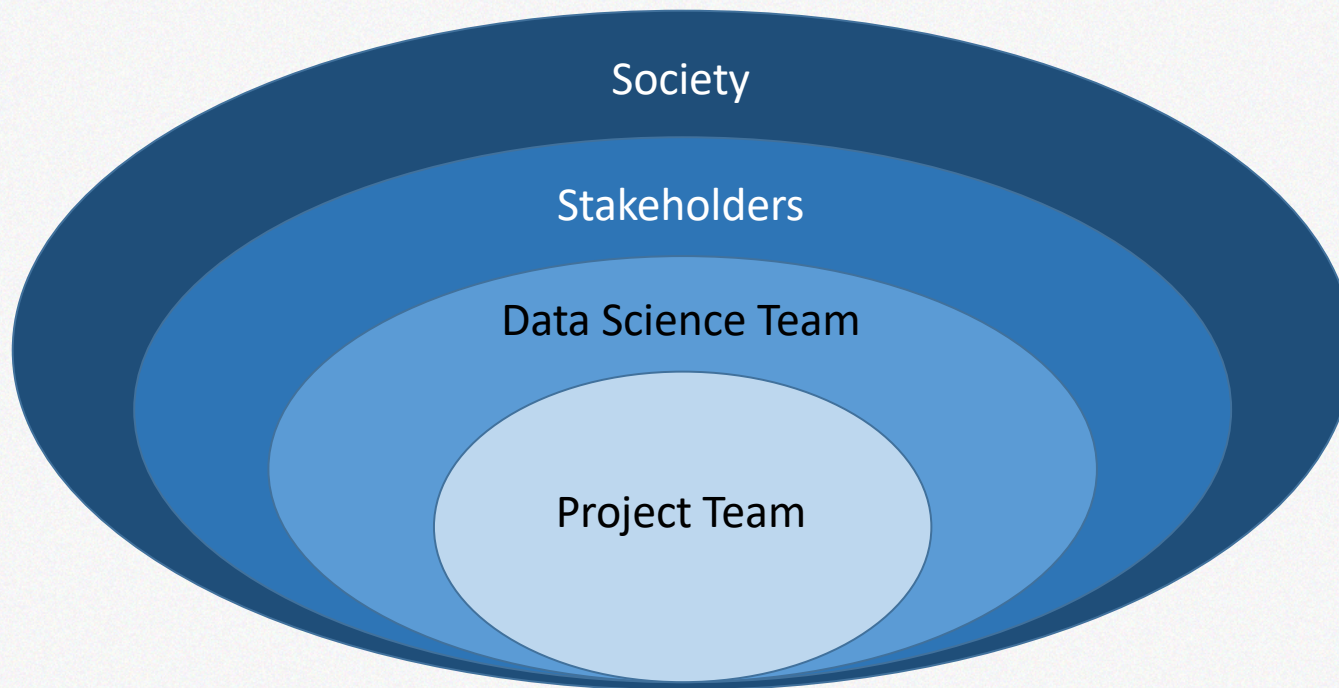
Discuss the importance to validate properly statistical models

Share techniques to validate different dimensions of a model

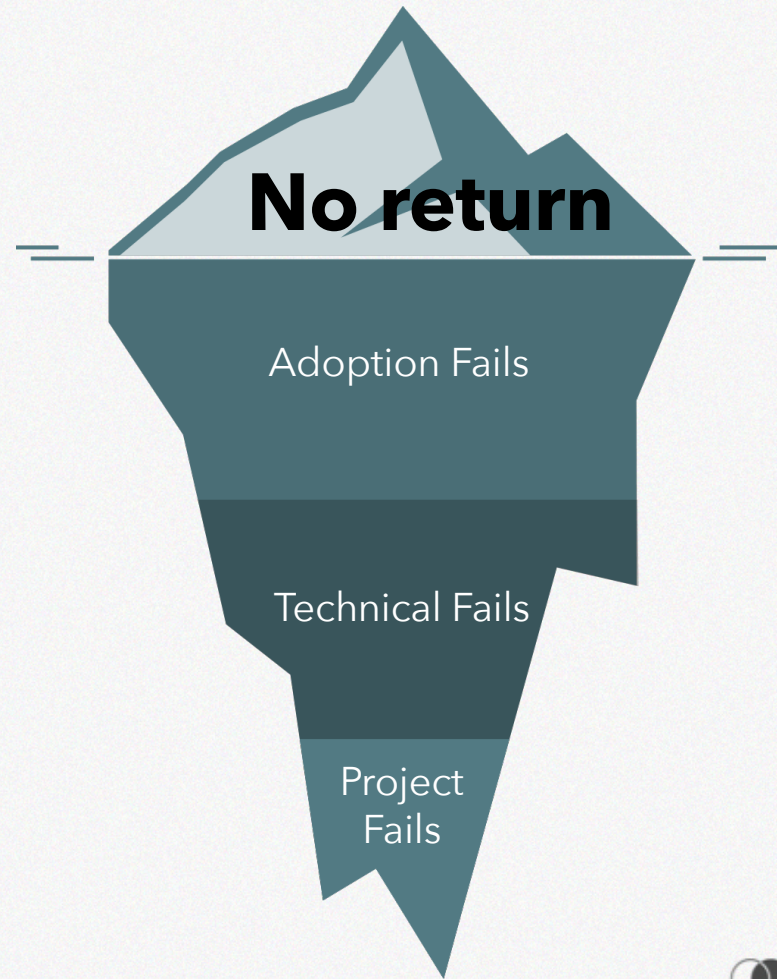
Demonstrate some of the techniques in a tutorial

**87% of machine learning proof
of concepts are never put in
production**

How Can Model Validation Help?

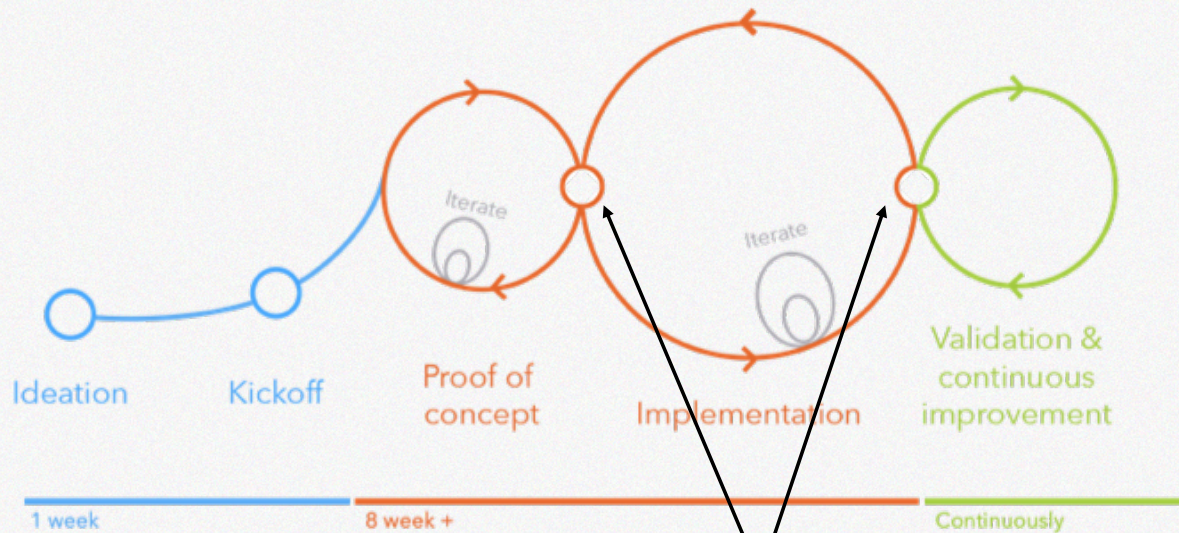


Iceberg of Machine Learning Failure



Project Development

Project Development



Project Validation

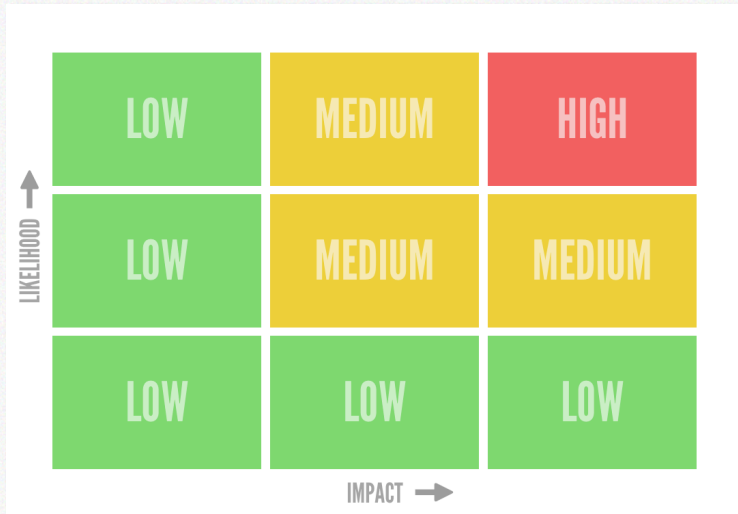
Define/ prevent project risks

Define/ prevent technical risks

Validation prior To deployment

On-going monitoring

Model Validation Framework



Wall between execution and validation

Identify unusual elements

Document and prioritise the elements

Validation required prior to deployment

Define Project Risks

Training data quality

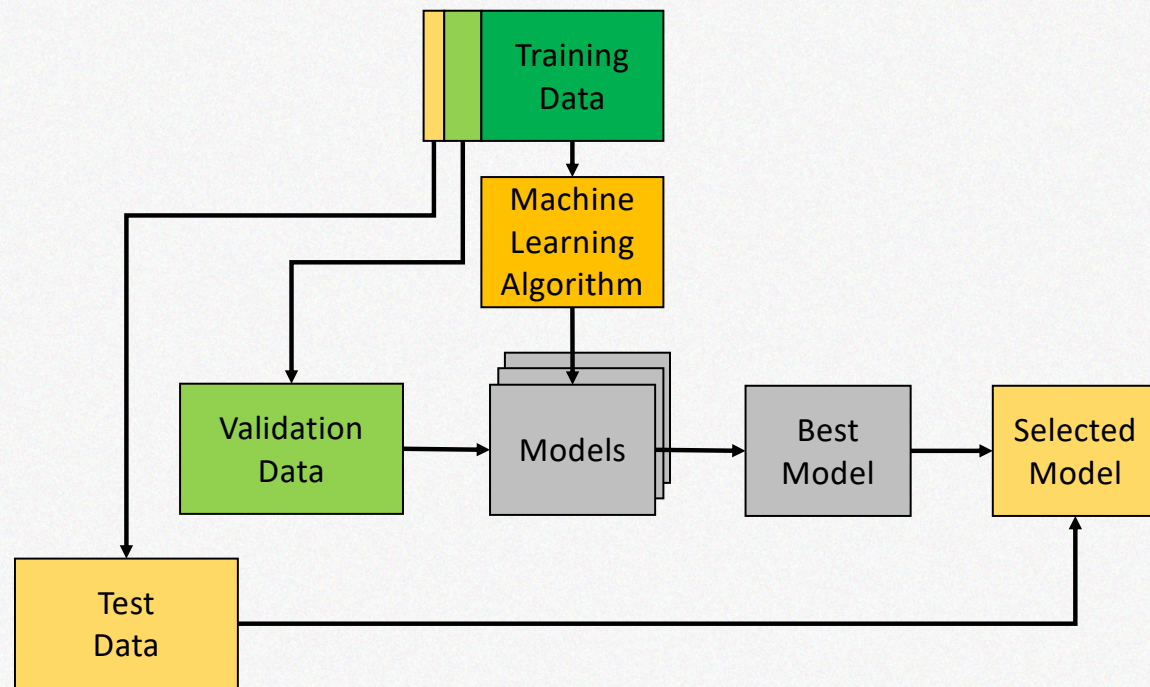
False positive and false negative biases

Production data accessibility

Integration to systems

User experience

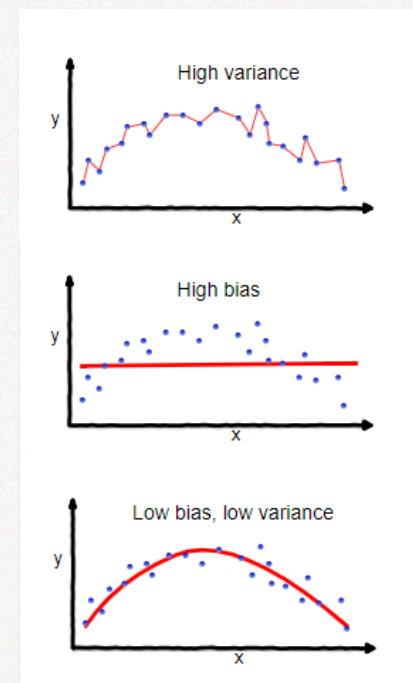
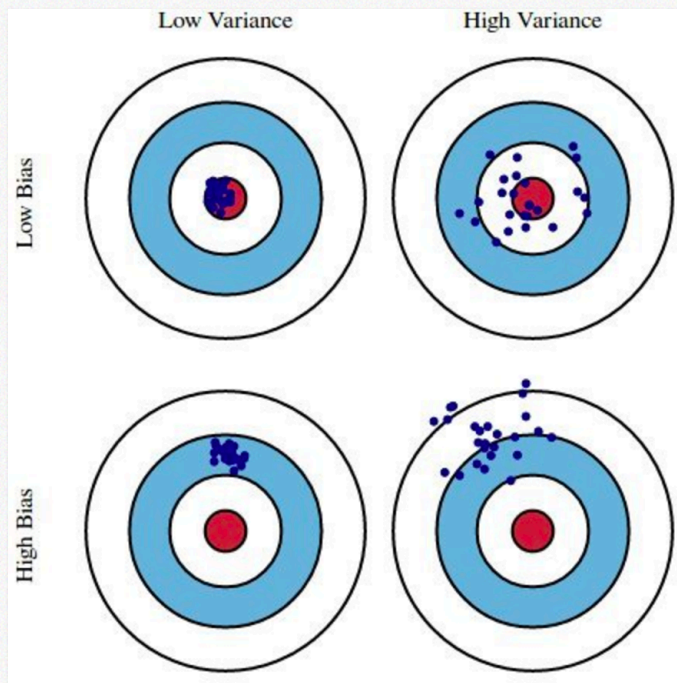
Model Validation Basics



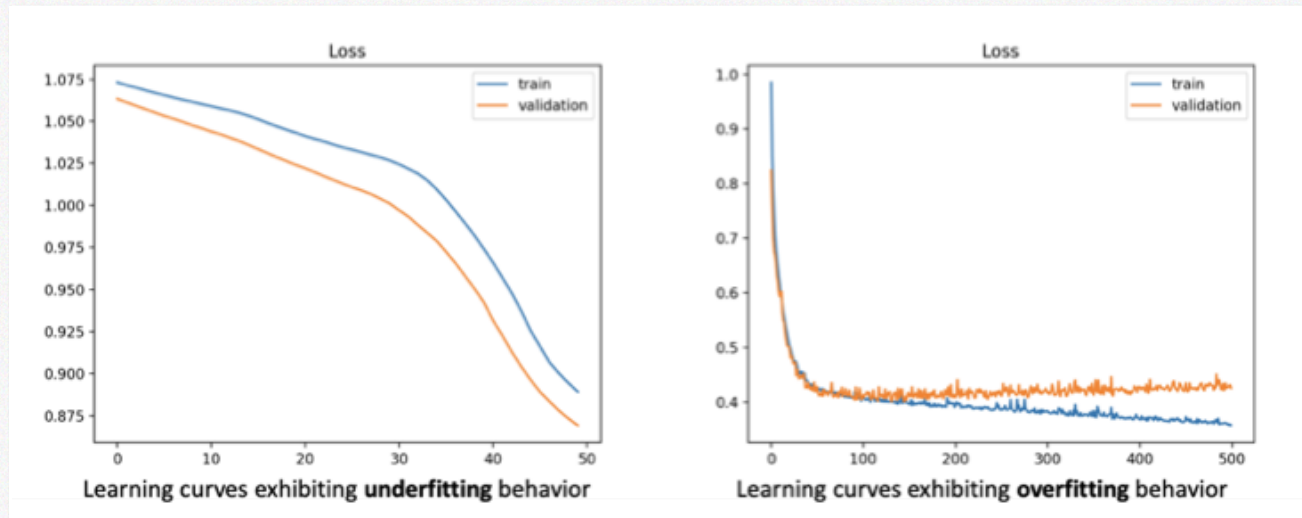
Things To Validate

Dimensions	Main Objective
Bias Error	Is the data useful?
Variance Error	Is the model robust?
Model Fit	Is the model predicting well new data?
Model Comparison	Is the new model better than simpler alternatives?
Discrimination Bias	Is the model too biased towards certain variables?
Adversarial Sensitivity	Is the model robust to extreme scenarios?

Bias vs. Variance

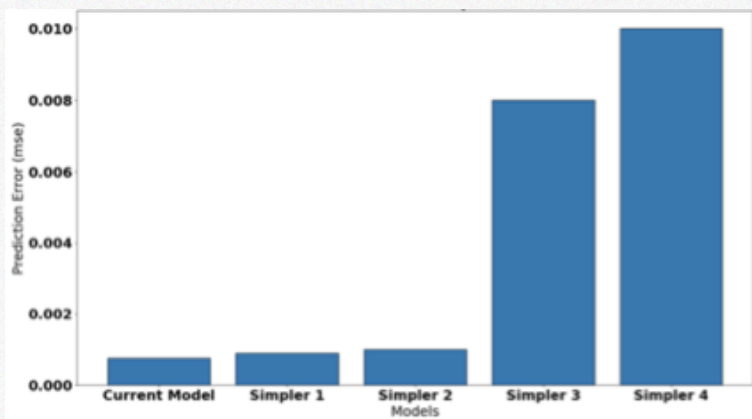


Model Fit Assessment



Model Comparison Assessment

Model Error comparison between models



It's not just about accuracy!

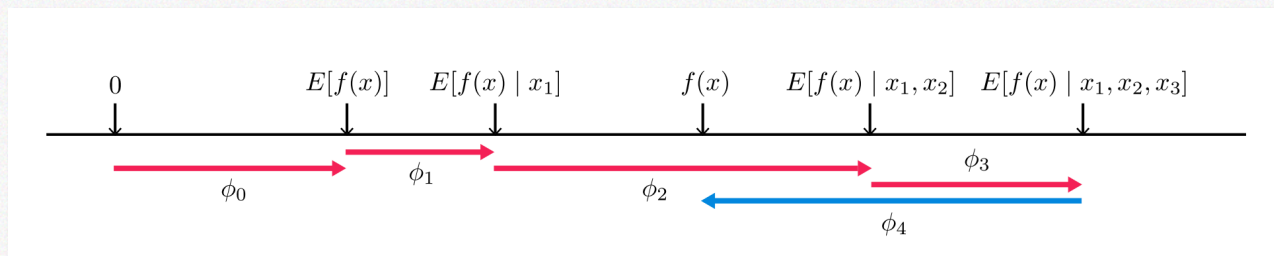
Other useful comparative to check:

- Model robustness
- Training time
- Inference time
- Storage needed

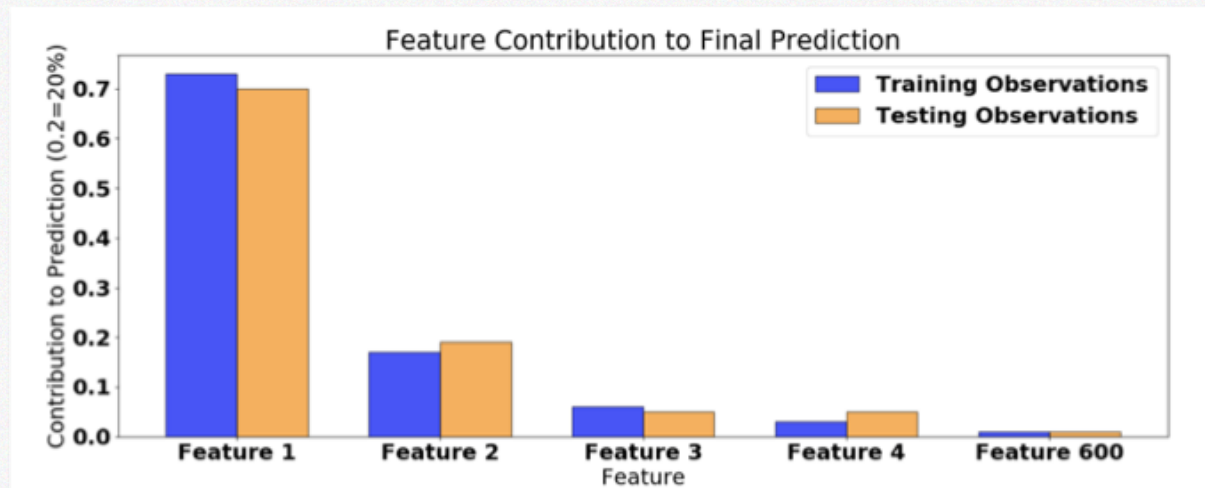
Discriminatory Features Assessment

SHAP, an agnostic interpretability technique is a good tool to assess the impact of the features on predictions.

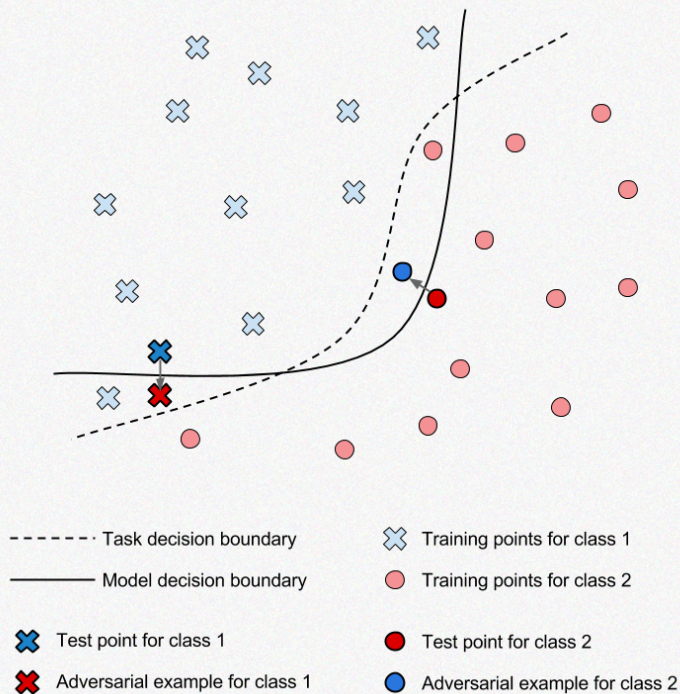
SHAP is able to get marginal importance per feature.



Discriminatory Features Assessment



Sensitivity Assessment



X

Original input
(correctly classified)

δ

X'

Original input + perturbation = adversarial example
(misclassified)

Adversarial examples differ based on the method used to craft the perturbation (e.g. Fast Gradient Sign Method, FGSM).

Sensitivity Assessment

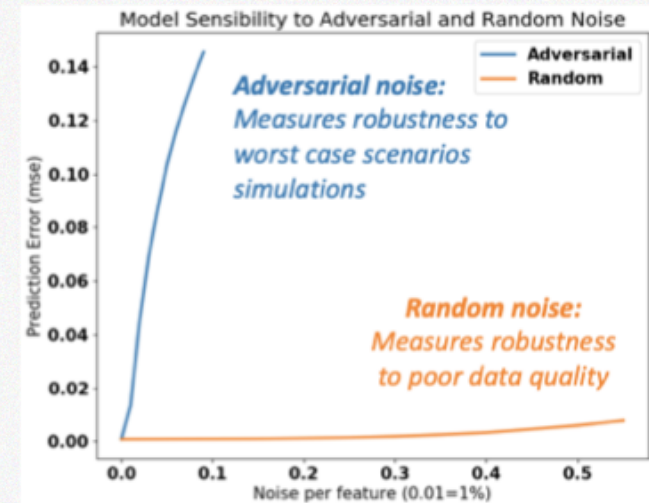
Noise is a slight modification in the input data

Random noise

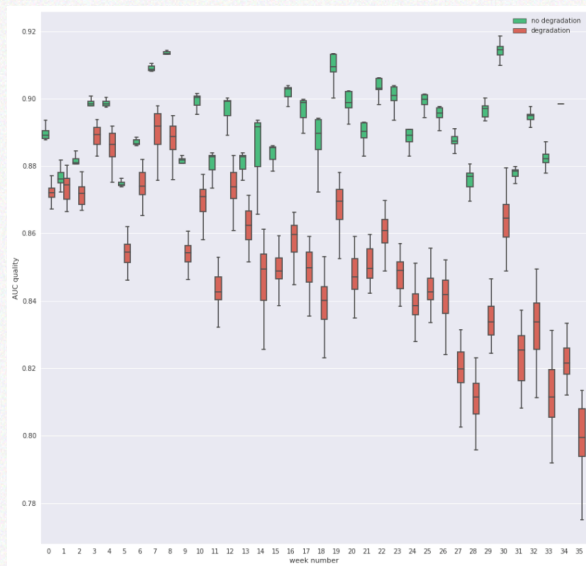
random noise can be added to each feature.

Adversarial noise

modifying each feature of the perfect amount and in the perfect direction to intentionally decrease the model's performance.



Monitor your Model



It is critical to monitor model's performance

Same assessments for initial validation using new data works fine

Look at prediction error increase for proactive maintenance

Tutorial

Conclusions

Models have to be validated in-depth before launch

Validation has to be more than simply about accuracy

Models have to be monitored frequently to avoid degradation

Thank you!