Institute
and Faculty
of Actuaries

Machine Learning: “Pricing” the Way Forward

Navarun Jain
Actuarial Analyst
Lux Actuaries & Consultants
Dubai

18 April 2019

Agenda

Machine Learning – The Concept

Gradient Boosters

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN's Work and Learn

What does it mean to "learn"?

- Gradient Descent

Applications to Insurance Data

Interpreting Machine Learning Models

- Measuring Feature Importance
- Finding Variable Interactions

Key Takeaways and Conclusions

Institute and Faculty of Actuaries

Machine Learning

18 April 2019

Agenda

Machine Learning – The Concept

Gradient Boosters

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN's Work and Learn

What does it mean to "learn"?

- Gradient Descent

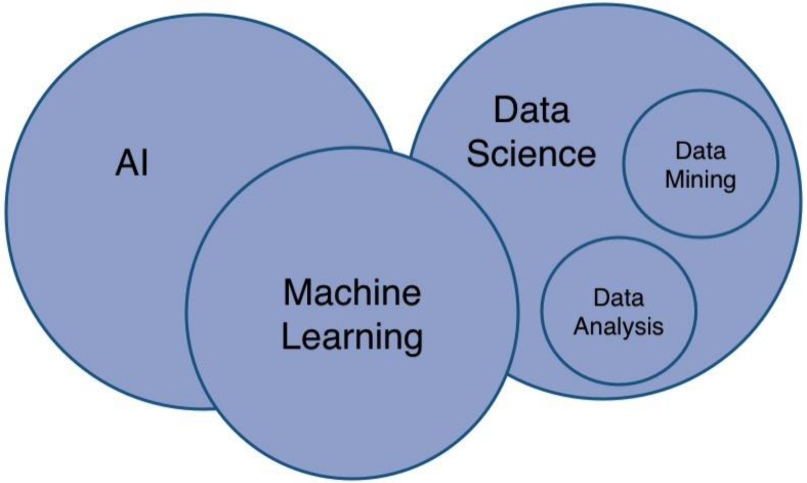
Applications to Insurance Data

Interpreting Machine Learning Models

- Measuring Feature Importance
- Finding Variable Interactions

Key Takeaways and Conclusions

Machine Learning



Machine Learning

The "teaching a kid math" analogy

Institute and Faculty of Actuaries

Machine Learning

All about
patterns!!!

The Roadmap

All about patterns!!!

Computer systems <u>learn</u>
from data

We **<u>train</u>** the system → System **<u>learns</u>** → Then performs operations **<u>on its own</u>**

Institute
and Faculty
of Actuaries

The Roadmap

All about patterns!!!

Computer systems learn from data

We **train** the system → System **learns** → Then performs operations **on its own**

Training phase 1: data is fed into the algorithm, relevant fields and records sorted from data to retrieve **active dataset**

Institute and Faculty of Actuaries

The Roadmap

All about patterns!!!

Computer systems learn from data

We **train** the system → System **learns** → Then performs operations **on its own**

Training phase 2: **Model Fitting** – algorithm decodes hidden patterns and relationships in the data

Institute and Faculty of Actuaries

The Roadmap

All about patterns!!!

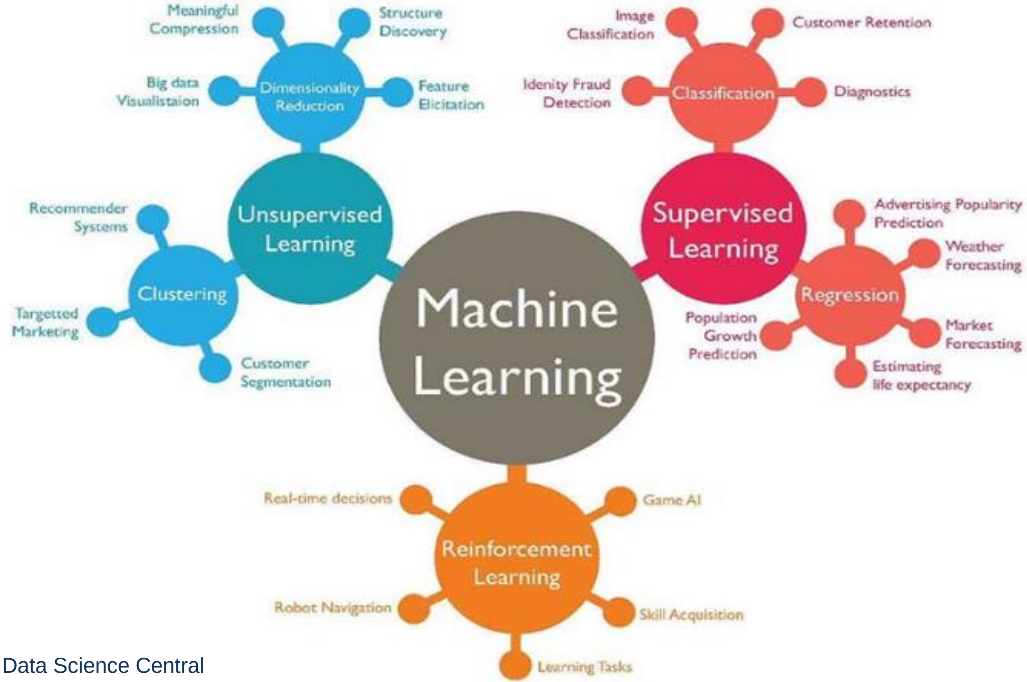
Computer systems learn
from data

We **train** the system → System **learns** → Then performs operations **on its own**

Testing phase: new data fed into system, algorithm uses patterns & relationships learnt during the training phase to predict new cases

Institute and Faculty of Actuaries

Types of Algorithms



Source: Data Science Central

Institute and Faculty of Actuaries

With ML, no need to…

With ML, no need to…

- …make assumptions about distributions

Institute and Faculty of Actuaries

With ML, no need to…

- …make assumptions about distributions
- …worry about possible correlations between predictors

Institute and Faculty of Actuaries

With ML, no need to…

- …make assumptions about distributions
- …worry about possible correlations between predictors
- …look for interactions between predictors

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Gradient Boosters

18 April 2019

Agenda

Machine Learning – The Concept

Gradient Boosters

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN's Work and Learn

What does it mean to "learn"?

- Gradient Descent

Applications to Insurance Data

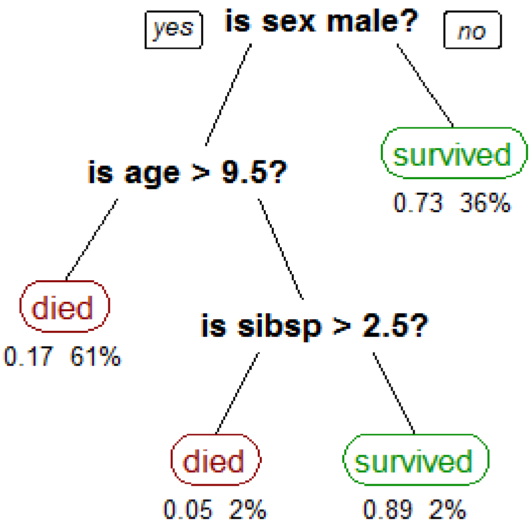
Interpreting Machine Learning Models

- Measuring Feature Importance
- Finding Variable Interactions

Key Takeaways and Conclusions

Institute and Faculty of Actuaries

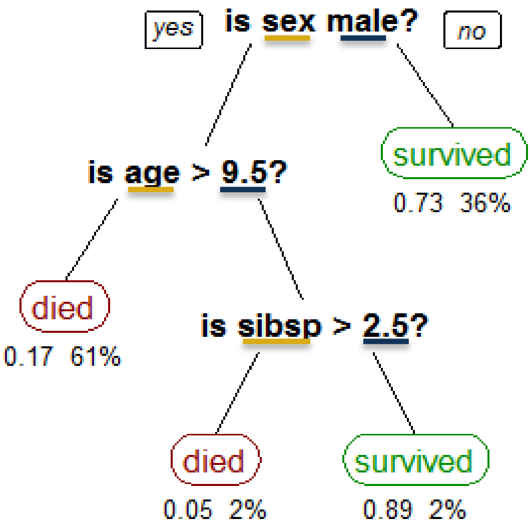
Decision Trees



Source: Wikipedia

Model is grown by recursively splitting the data into **decision boundaries** using the **feature space**

Decision Trees



Model is grown by recursively splitting the data into **decision boundaries** using the **feature space**

Source: Wikipedia

Institute and Faculty of Actuaries

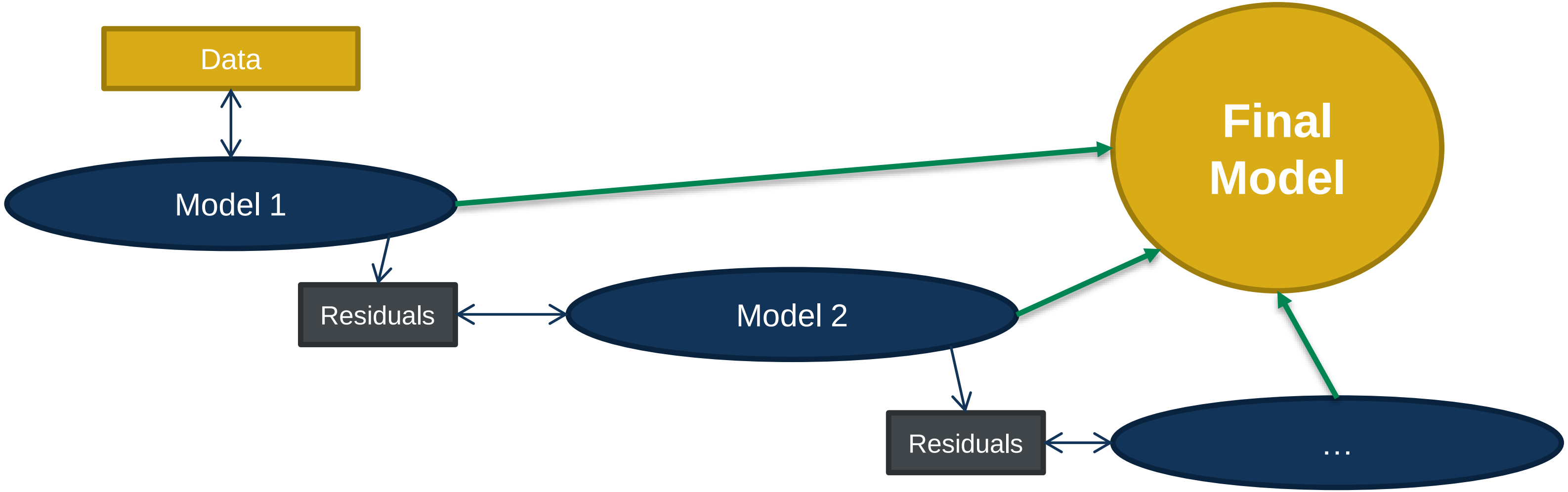
Boosting

- Converts weak learners into a single strong learner by aggregating them

Institute and Faculty of Actuaries

Boosting

- Converts weak learners into a single strong learner by aggregating them



Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Artificial Neural Networks

Making computers think like we do!

18 April 2019

Agenda

Machine Learning – The Concept

Extreme Gradient Boosting

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN's Work and Learn

What does it mean to "learn"?

- Gradient Descent

Applications to Insurance Data

Interpreting Machine Learning Models

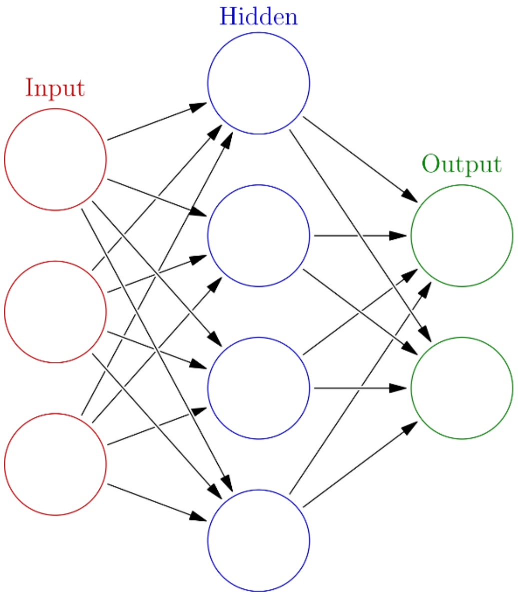
- Measuring Feature Importance
- Finding Variable Interactions

Key Takeaways and Conclusions

Institute and Faculty of Actuaries

Artificial Neural Networks

Structured Sequential model



Structured: A Neural Network has a defined structure that consists of 3 types of layers

Sequential: Information flows in a sequence from one layer to the next, undergoing operations at each layer – almost like an assembly line

Institute and Faculty of Actuaries

How ANN's Work

How ANN's Work

- Data in every neuron is transformed by an activation function:

$$h_k(x) = g(\beta_{0k} + \sum_{i=1}^{n} x_i \beta_{ik})$$

$h_k(x)$ – k^{th} neuron in a hidden layer
β_{ik} - coefficient of the i^{th} previous-layer neuron on above neuron

How ANN's Work

- Data in every neuron is transformed by an activation function:

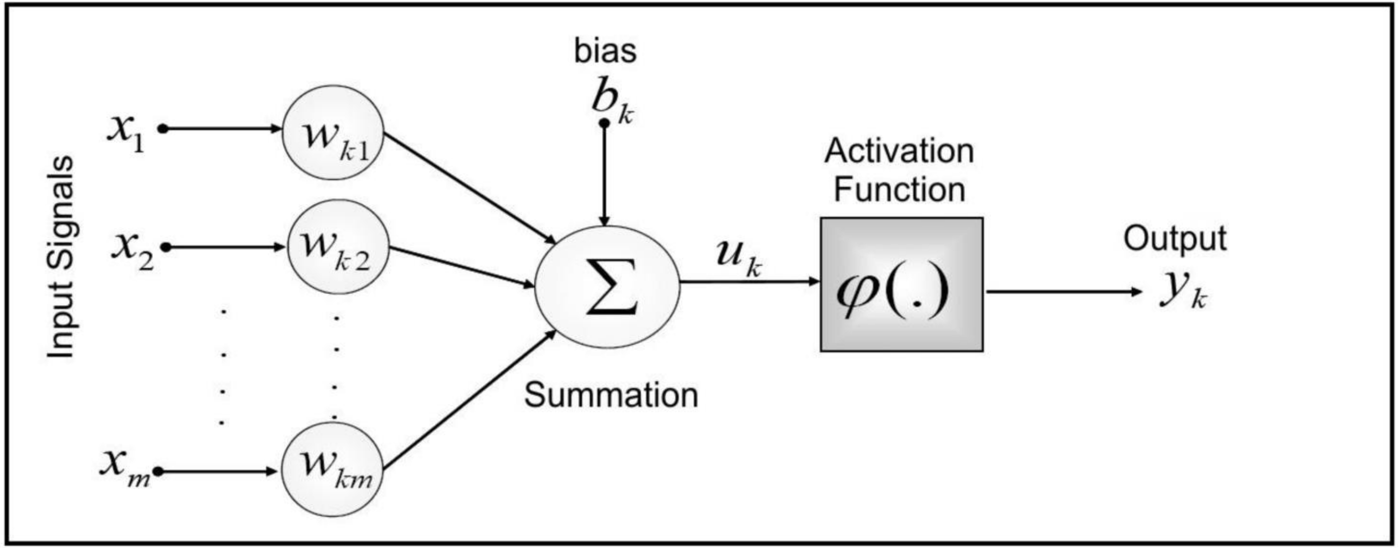
$$h_k(x) = g(\beta_{0k} + \sum_{i=1}^{n} x_i \beta_{ik})$$

$h_k(x)$ – k^{th} neuron in a hidden layer
β_{ik} - coefficient of the i^{th} previous-layer neuron on above neuron

- Activation function transforms the linear combination of inputs from one layer and sends it to the next layer.

Institute and Faculty of Actuaries

How ANN's Work

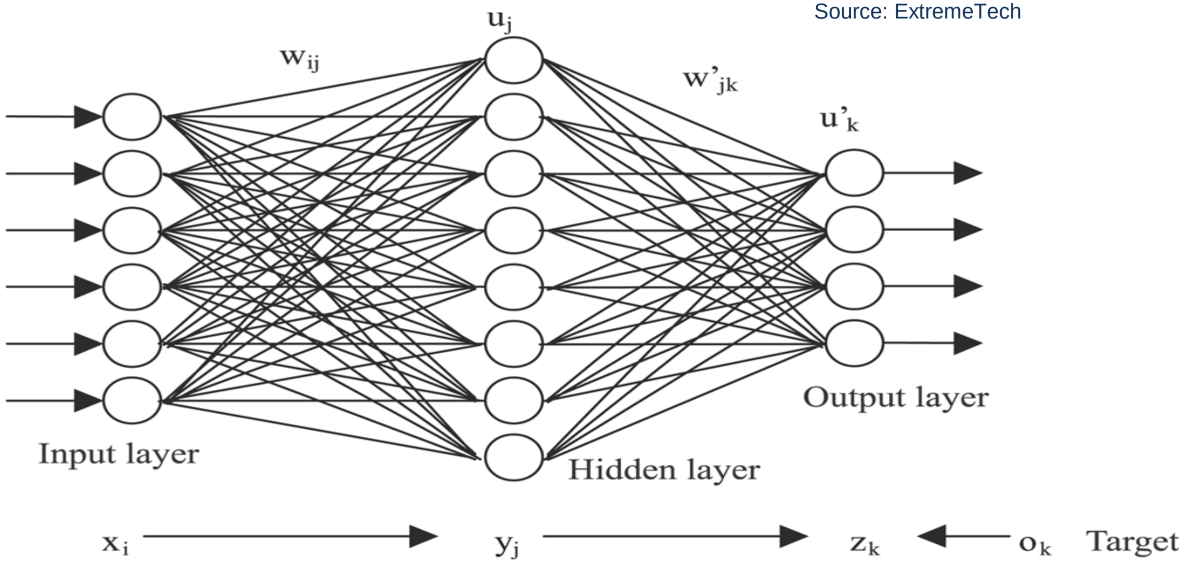


Source: MDPI

Institute
and Faculty
of Actuaries

How ANN's Work

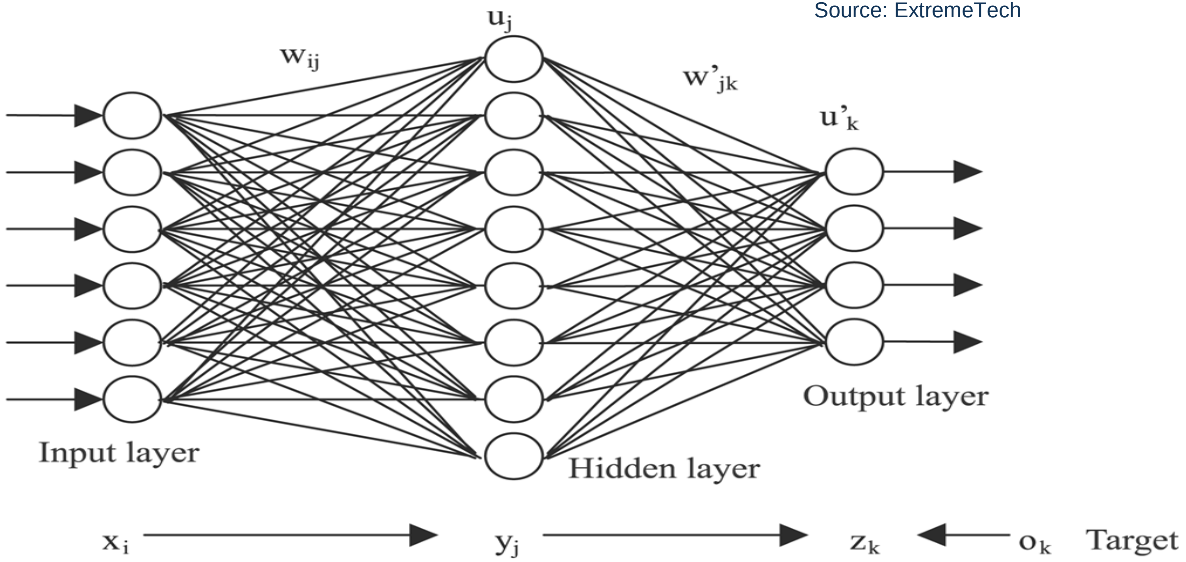
Source: ExtremeTech



Institute and Faculty of Actuaries

How ANN's Work

Source: ExtremeTech

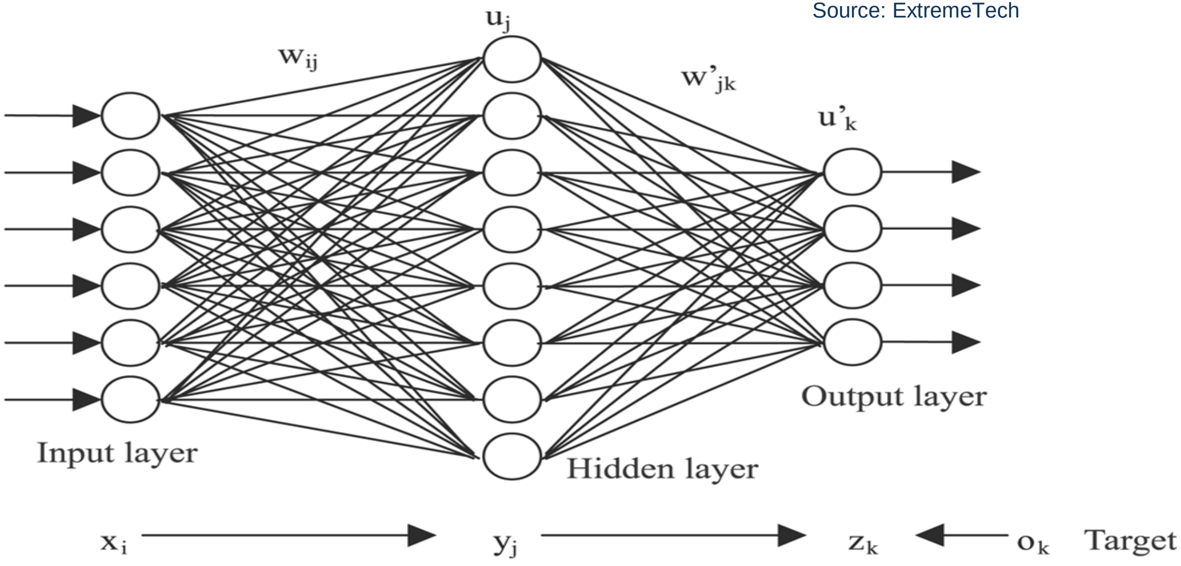


At first, each neuron is randomly assigned a weight – this measures the contribution of that neuron to the next layer

Institute and Faculty of Actuaries

How ANN's Work

Source: ExtremeTech

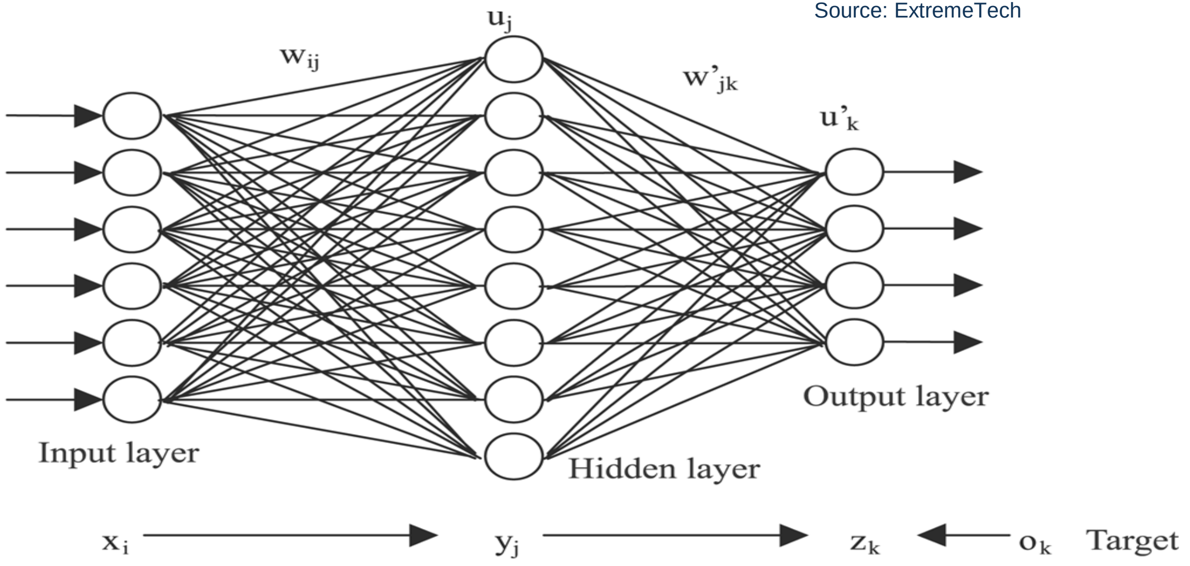


Data flows through network, predicted values calculated

Institute and Faculty of Actuaries

How ANN's Work

Source: ExtremeTech

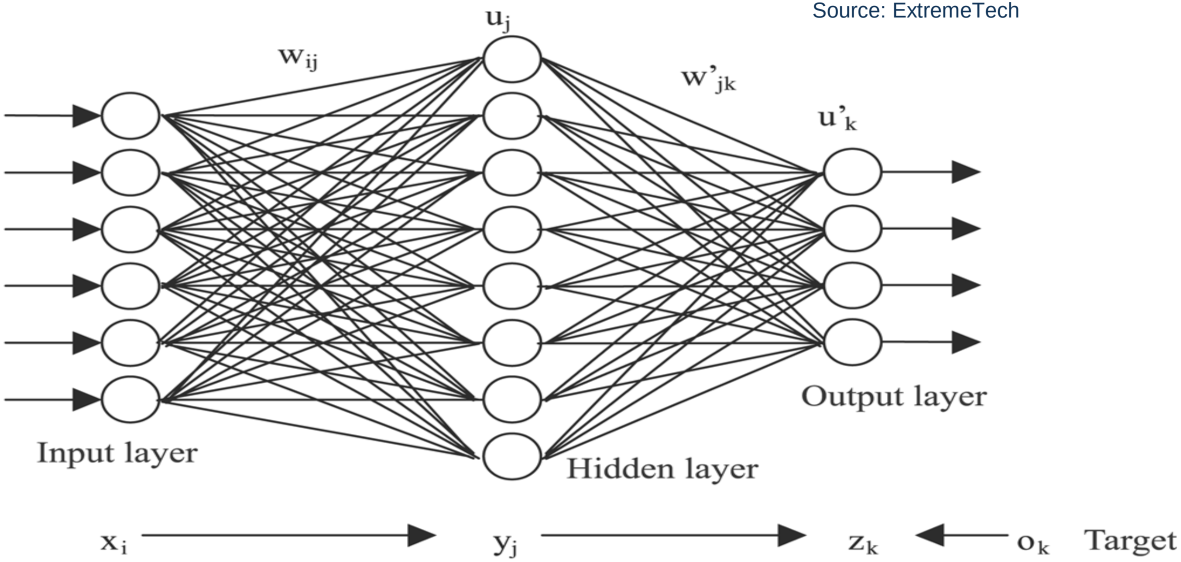


Predictions are compared with actuals based on a loss function

Institute and Faculty of Actuaries

How ANN's Work

Source: ExtremeTech



Weights are updated to reduce value of loss function

Institute and Faculty of Actuaries

What does it mean to “learn”?

Gradient Descent

18 April 2019

Agenda

Machine Learning – The Concept

Extreme Gradient Boosting

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN's Work and Learn

What does it mean to "learn"?

- Gradient Descent

Applications to Insurance Data

Interpreting Machine Learning Models

- Measuring Feature Importance
- Finding Variable Interactions

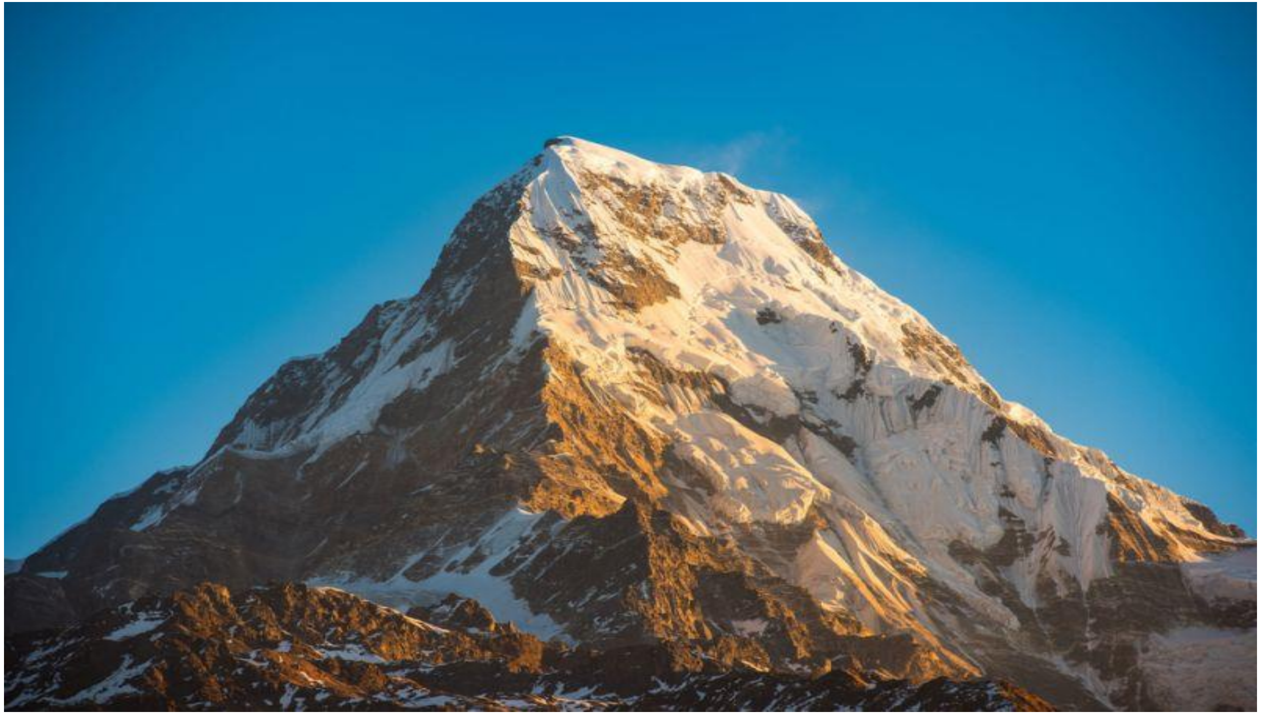
Key Takeaways and Conclusions

Institute and Faculty of Actuaries

Gradient Descent

Institute and Faculty of Actuaries

Gradient Descent



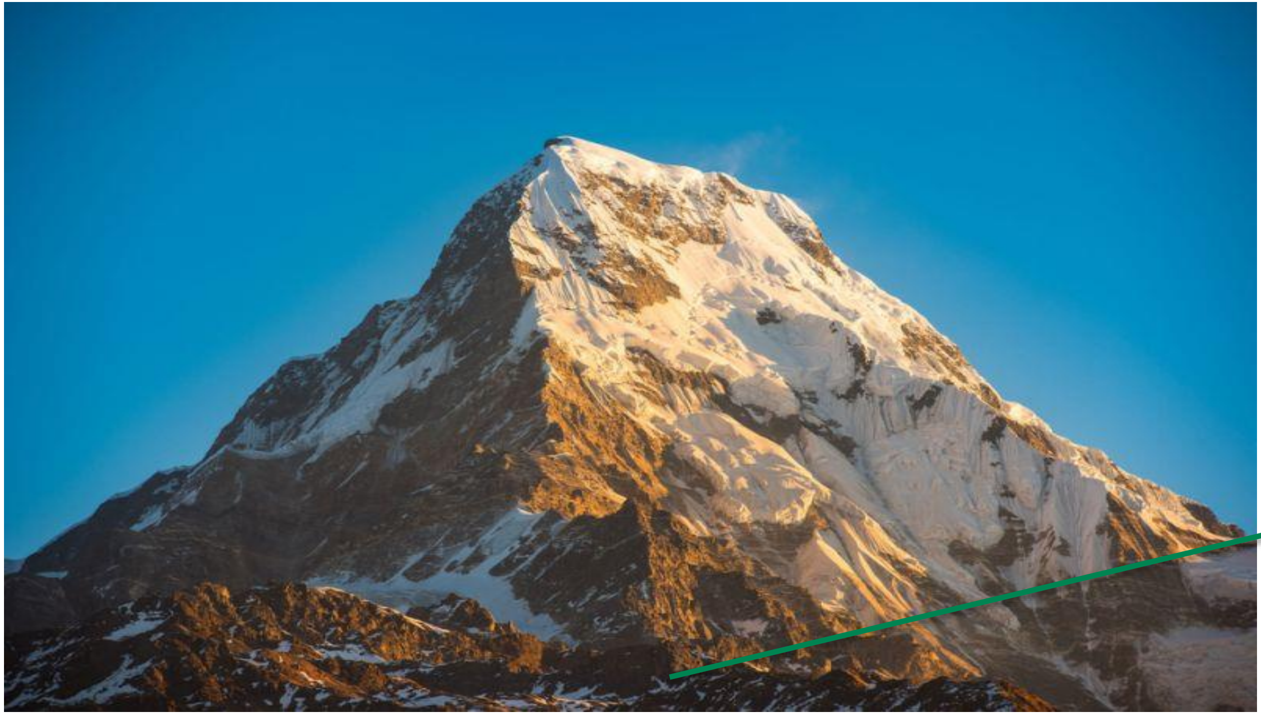
Institute and Faculty of Actuaries

Gradient Descent

We are here

Institute and Faculty of Actuaries

Gradient Descent



Want to go there

Institute and Faculty of Actuaries

Gradient Descent

- Modelling continues until the following is minimized:

Gradient Descent

- Modelling continues until the following is minimized:

$$\nabla_{\mathrm{W}}\mathrm{L} = \frac{\delta L}{\delta W}$$

Gradient of the Loss function – measures change in loss function as model weights change

Institute and Faculty of Actuaries

Gradient Descent

- Modelling continues until the following is minimized:

$$\nabla_W L = \frac{\delta L}{\delta W}$$

Gradient of the Loss function – measures change in loss function as model weights change

- The above function is computed and a step is taken in the direction where it is minimized the most **relative to our current position**

Institute and Faculty of Actuaries

Gradient Descent

- Modelling continues until the following is minimized:

$$\nabla_W \mathrm{L} = \frac{\delta L}{\delta W}$$

Gradient of the Loss function – measures change in loss function as model weights change

- The above function is computed and a step is taken in the direction where it is minimized the most **relative to our current position**
- Size of this step is the **learning rate**

Institute and Faculty of Actuaries

Optimizing Neural Networks with GD

- Suppose for Neuron A and iteration t, the weight was found to be $W_{A(t)}$

Institute and Faculty of Actuaries

Optimizing Neural Networks with GD

- Suppose for Neuron A and iteration *t*, the weight was found to be $W_{A(t)}$
- Then, for iteration $t + 1$, weight is optimized to:

$$W_{A(t+1)} = W_{A(t)} - \eta \nabla_{W_{A(t)}} L$$

- $\boldsymbol{\eta}$ – Learning Rate Multiplier
- $\boldsymbol{\nabla_{W_{A(t)}} L}$ – Gradient of Loss Function w.r.t. weight of Neuron A at iteration *t*

Institute and Faculty of Actuaries

Optimizing Neural Networks with GD

- **Vanilla approach: Compute gradient for entire training sample and update weights based on that**

Institute and Faculty of Actuaries

Optimizing Neural Networks with GD

- **Vanilla approach: Compute gradient for entire training sample and update weights based on that**
  - No method to check if full convergence is achieved
  - What if different parameters work differently and require different optimization rates?

Institute and Faculty of Actuaries

Optimizing Neural Networks with GD

- **Vanilla approach: Compute gradient for entire training sample and update weights based on that**
  - No method to check if full convergence is achieved
  - What if different parameters work differently and require different optimization rates?
- **Stochastic Gradient Descent: Compute gradient for each individual point in the training sample and update weights iteratively for every sample**

Institute and Faculty of Actuaries

Optimizing Neural Networks with GD

- **Vanilla approach: Compute gradient for entire training sample and update weights based on that**
  - No method to check if full convergence is achieved
  - What if different parameters work differently and require different optimization rates?
- **Stochastic Gradient Descent: Compute gradient for each individual point in the training sample and update weights iteratively for every sample**
  - Too slow – Might cause algorithm to crash or give up for extremely large datasets, thus potentially preventing full convergence

Institute and Faculty of Actuaries

Adaptive Learning - RMSProp

Institute and Faculty of Actuaries

Adaptive Learning - RMSProp

- Different parameters may have different gradients

Institute and Faculty of Actuaries

Adaptive Learning - RMSProp

- Different parameters may have different gradients
- For each weight, RMSProp computes the moving average of its squared gradients

Institute and Faculty of Actuaries

Adaptive Learning - RMSProp

- Different parameters may have different gradients
- For each weight, RMSProp computes the moving average of its squared gradients
- Current gradient is divided by the square root of this average

$$E[g^2]_t = \beta E[g^2]_{t-1} + (1 - \beta)(\nabla_{W_{A(t)}} L)^2$$

$$W_{A(t+1)} = W_{A(t)} - \frac{\eta}{\sqrt{E[g^2]_t}} \nabla_{W_{A(t)}} L$$

- $\boldsymbol{\beta}$ – Moving Average Parameter (0.9 is a good value)
- $\boldsymbol{g}$ – Gradient of Loss function

Institute and Faculty of Actuaries

Institute
and Faculty
of Actuaries

Applications to Insurance Data

dataCar from R's insuranceData package

18 April 2019

Agenda

Machine Learning – The Concept

Extreme Gradient Boosting

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN's Work and Learn

What does it mean to "learn"?

- Gradient Descent

Applications to Insurance Data

Interpreting Machine Learning Models

- Measuring Feature Importance
- Finding Variable Interactions

Key Takeaways and Conclusions

Institute and Faculty of Actuaries

Data Description

- Policyholder-level information on one-year vehicle insurance policies
- 67,856 records with following rating factors –
  - Vehicle value in $10,000’s
  - Vehicle body type (eg. Sedan, convertible, hatchback, bus & other levels)
  - Vehicle age (Levels 1-4 w/1 being the newest & 4 being the oldest)
  - Gender of driver
  - Area
  - Driver age category (Levels 1-6 w/1 being youngest & 6 being oldest)

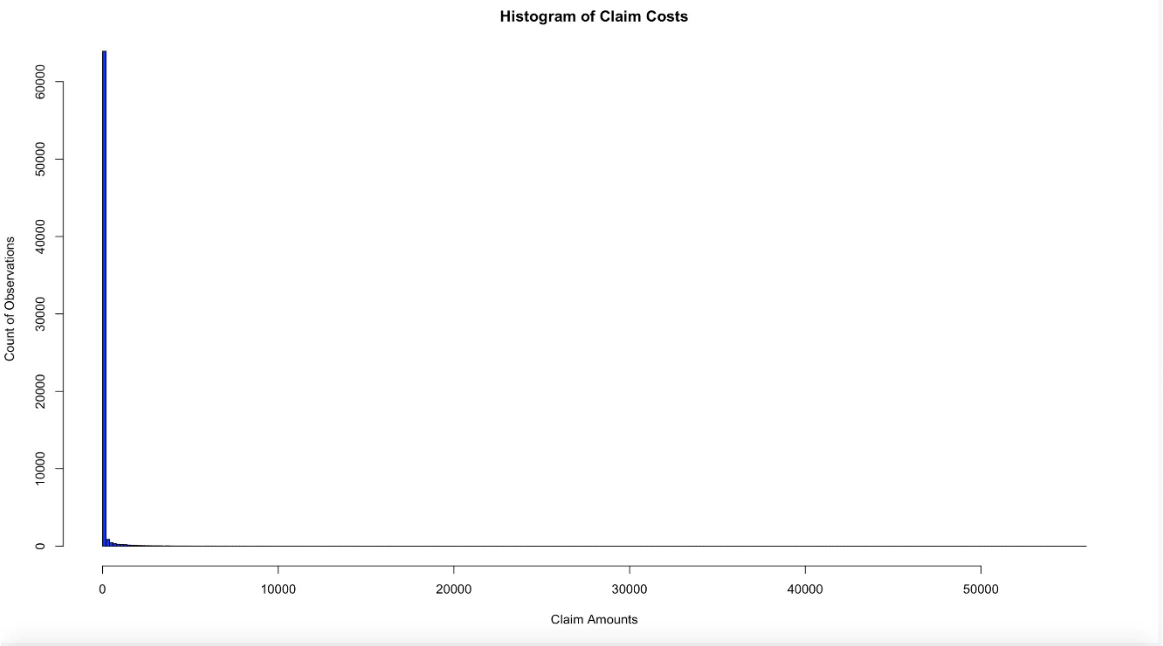
Institute and Faculty of Actuaries

Distribution of Claims

- Heavily skewed w/no-claim percentage of 93.2%

Distribution of Claims

- Heavily skewed w/no-claim percentage of 93.2%

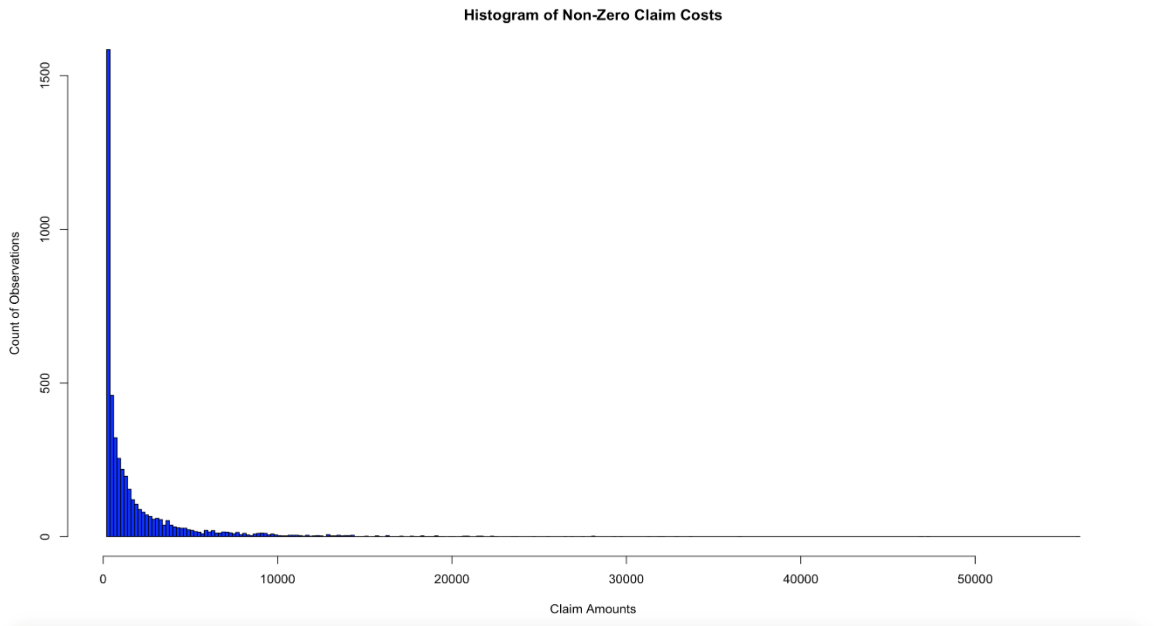


Distribution of raw claims data

Institute and Faculty of Actuaries

Distribution of Claims

- Heavily skewed w/no-claim percentage of 93.2%

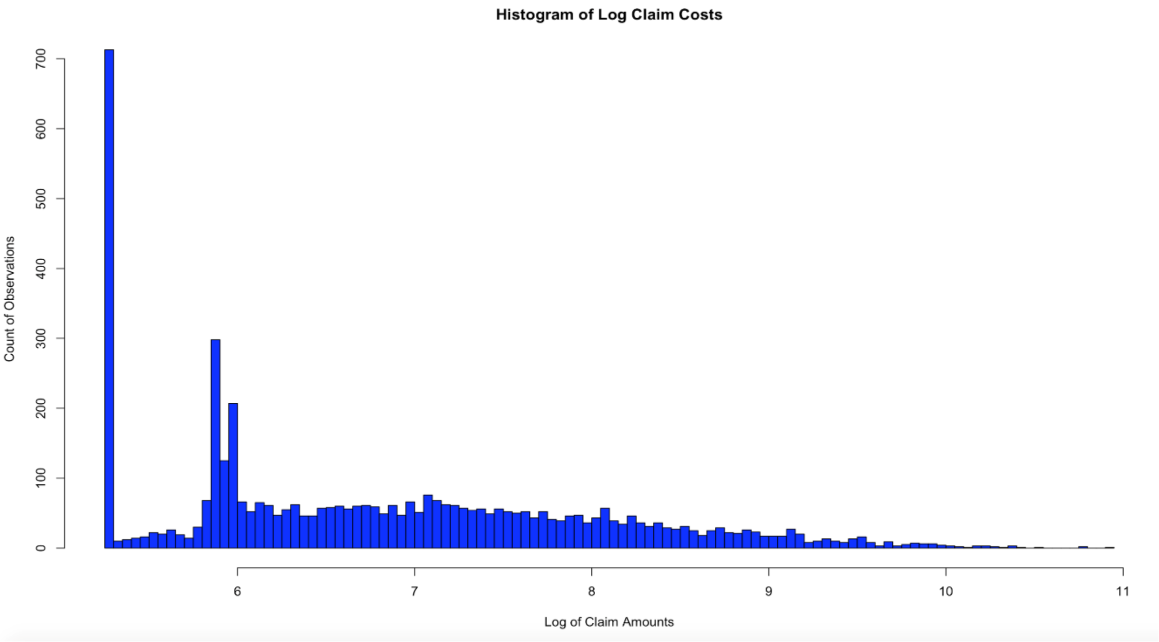


Distribution of
non-zero claims only

Institute and Faculty of Actuaries

Distribution of Claims

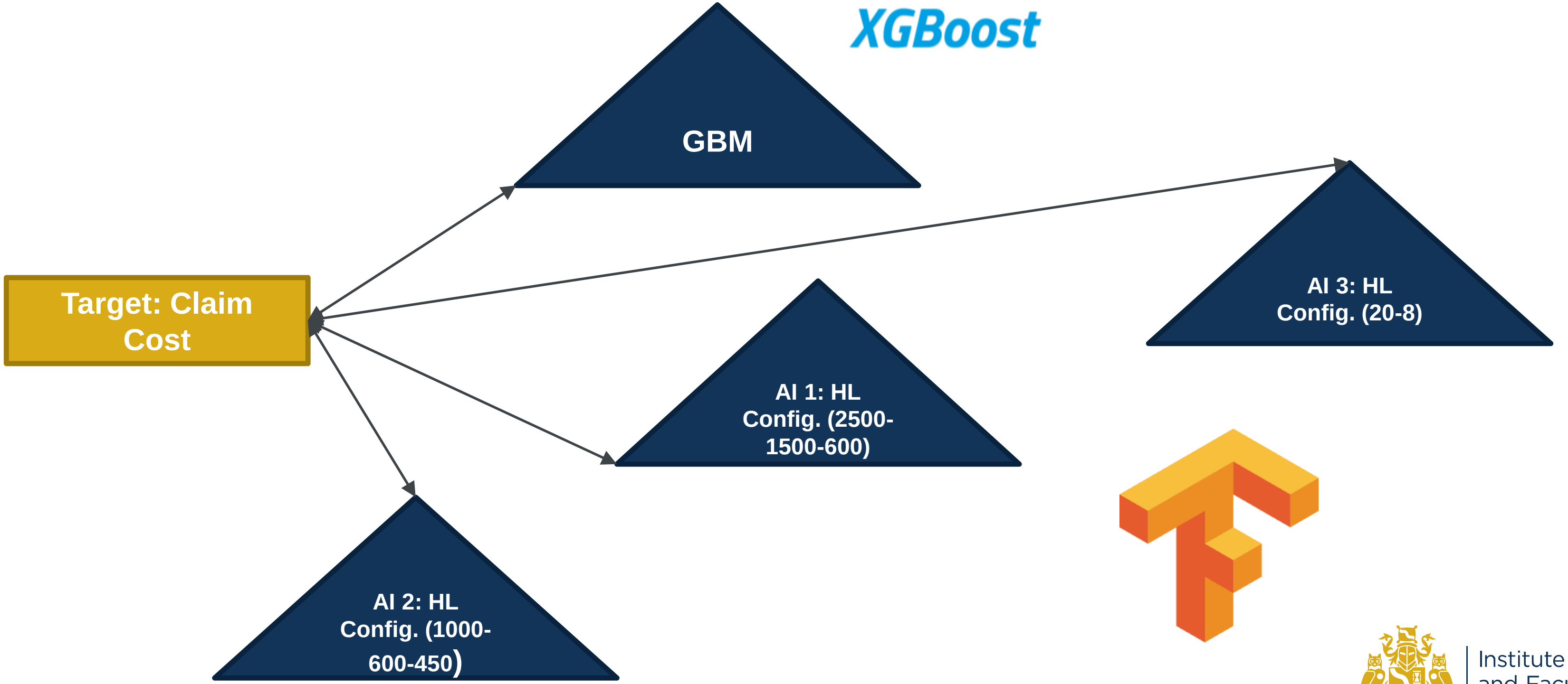
- Heavily skewed w/no-claim percentage of 93.2%



Distribution of the logarithm of claims

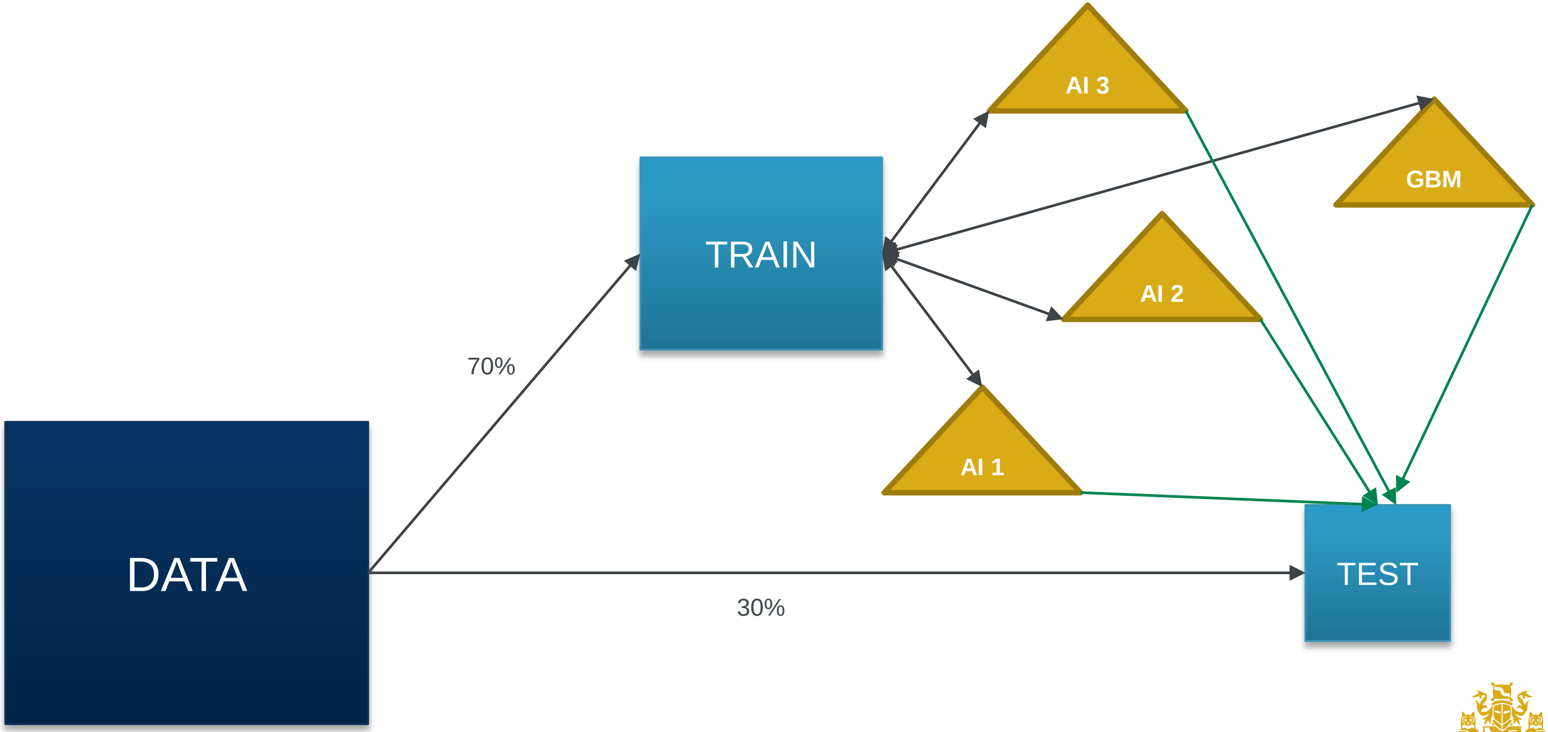
Institute and Faculty of Actuaries

Models


XGBoost
GBM
Target: Claim Cost
AI 3: HL Config. (20-8)
AI 1: HL Config. (2500-1500-600)
AI 2: HL Config. (1000-600-450)

Approach 1 – Standard Fit

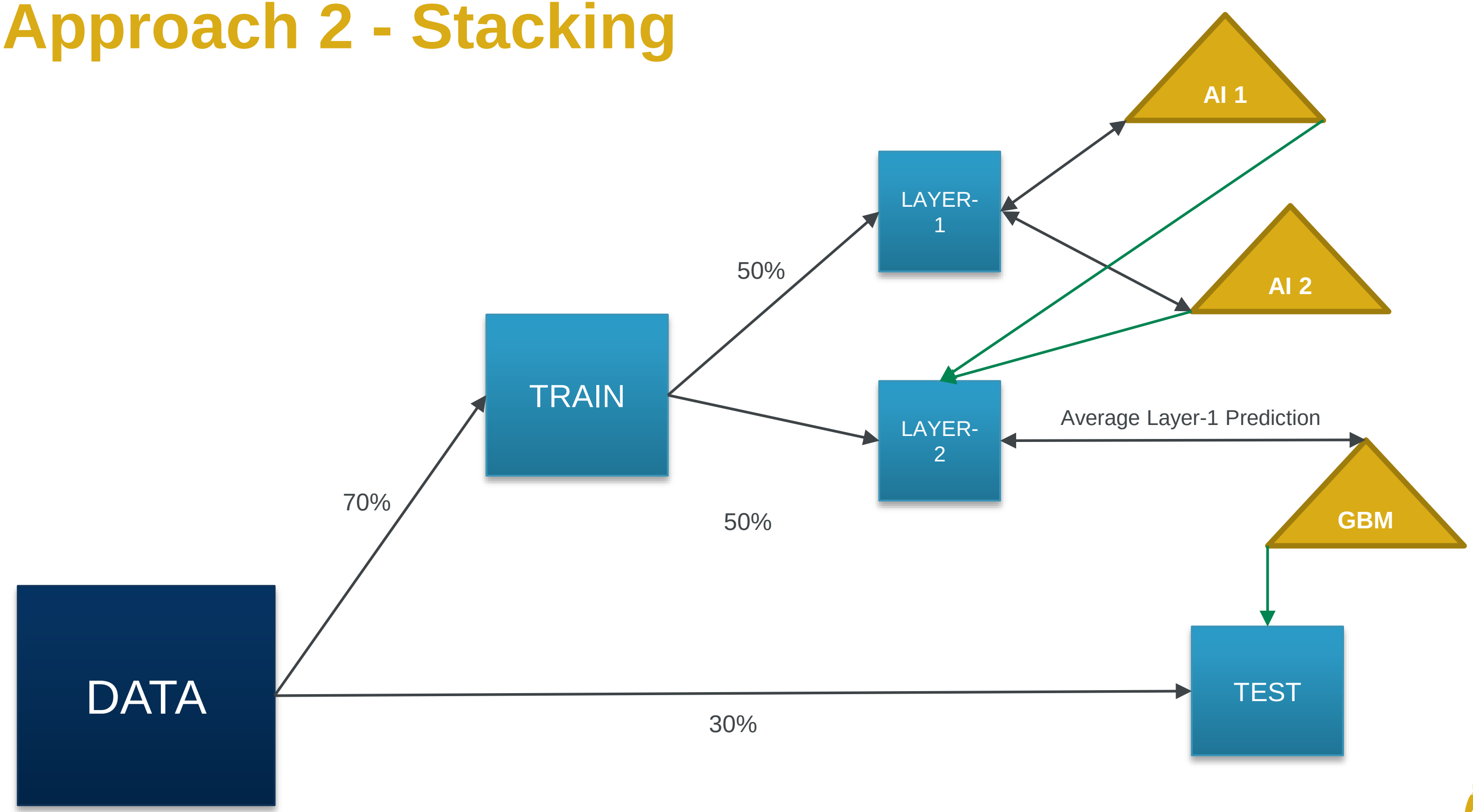
Approach 1 – Standard Fit



Institute and Faculty of Actuaries

Approach 2 - Stacking

Approach 2 - Stacking



Institute and Faculty of Actuaries

Model Comparison

Model Comparison

Model	Test RMSE ($\times\ 10^2$)	Test MAE ($\times\ 10^2$)
Tweedie GLM	9.51	2.702
GBM	10.43	2.168
AI 1: HLC (2500-1500-600)	15.01	3.614
AI 2: HLC (1000-600-450)	14.02	3.641
AI 3: HLC (20-8)	11.89	8.814
Average	11.28	3.112
Stack Model (Approach 2)	9.94	2.387

Institute and Faculty of Actuaries

Interpreting Machine Learning Models

18 April 2019

Agenda

Machine Learning – The Concept

Extreme Gradient Boosting

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN’s Work and Learn

What does it mean to “learn”?

- Gradient Descent

Applications to Insurance Data

Interpreting Machine Learning Models

- Measuring Feature Importance
- Finding Variable Interactions

Key Takeaways and Conclusions

Institute and Faculty of Actuaries

Types of Interpretation

Types of Interpretation

Interpretability

GLOBAL

Trying to understand the predictions on an *overall* level ***– In general, why does a model behave the way it does?***

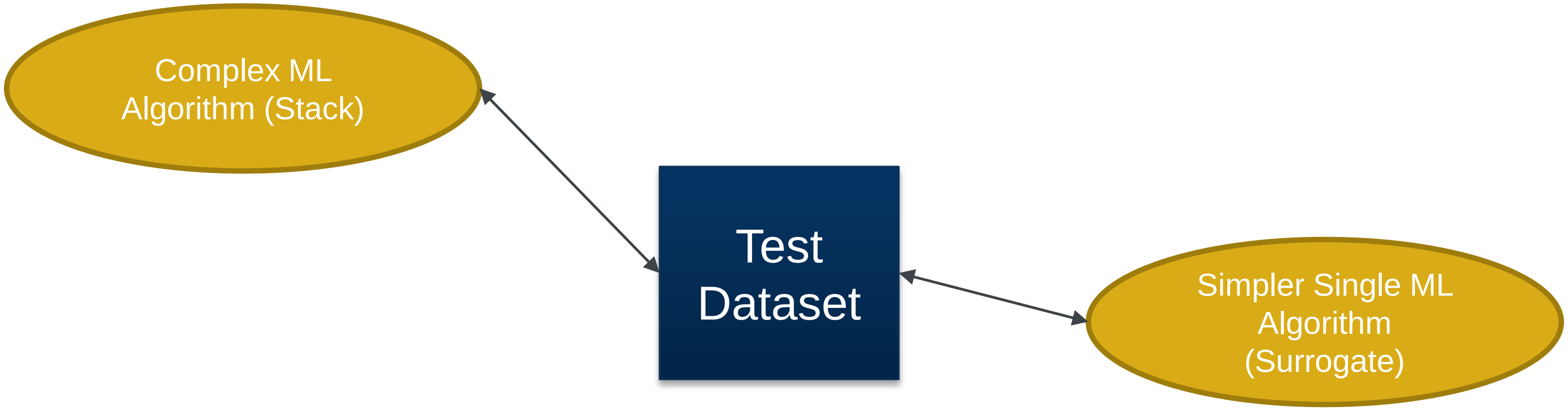
LOCAL

Trying to understand predictions for *specific records* – ***For a given record, what led the model to predict what it did?***

Institute and Faculty of Actuaries

Step 1: Building a Surrogate

Step 1: Building a Surrogate



Since the Stack isn't a model by itself, approximate it using a robust model

Institute and Faculty of Actuaries

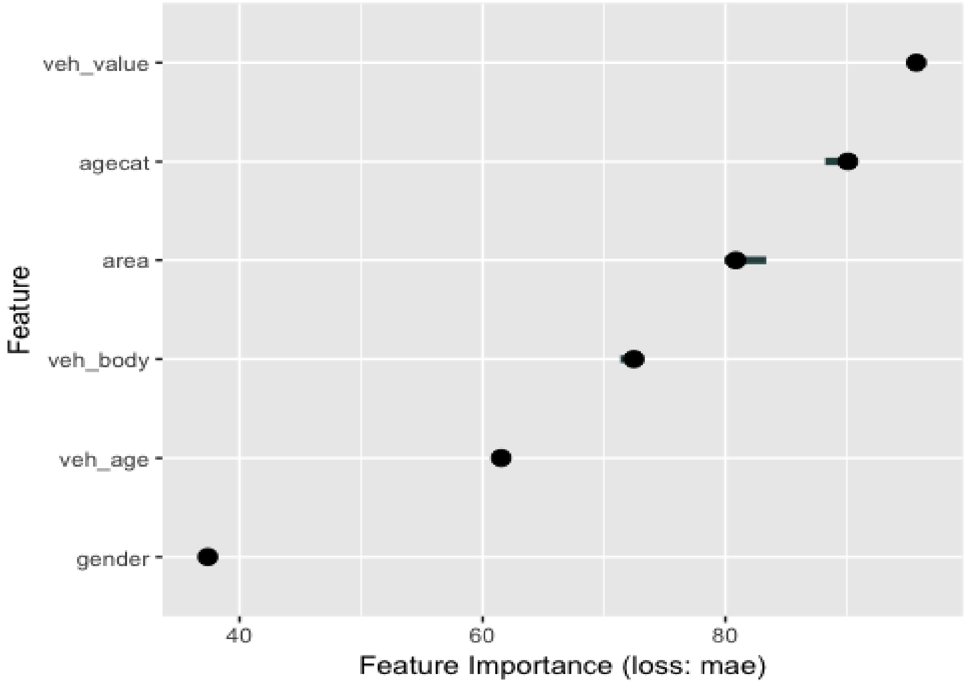
Global Interpretation

Global Interpretation

- Feature Importance
- Interaction Effects

Institute and Faculty of Actuaries

Feature Importance - dataCar



Vehicle Value, Driver Age and Geographical Location seem to be the key drivers of claims

Institute and Faculty of Actuaries

Interaction Effects

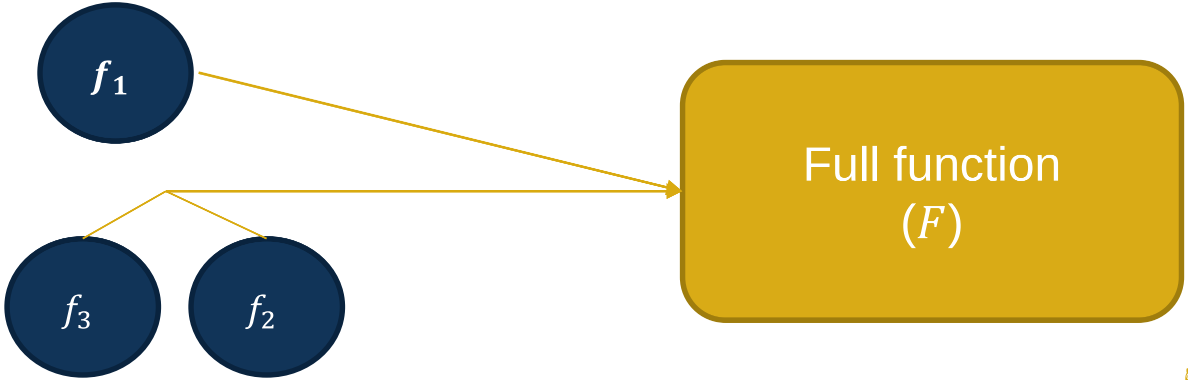
Interaction Effects

- For a feature f, algorithm computes partial function only dependent on f and partial function solely dependent on each of the other features
- If variance of full (true) function can be fully explained by the sum of the above partials, no interaction is attributed to f

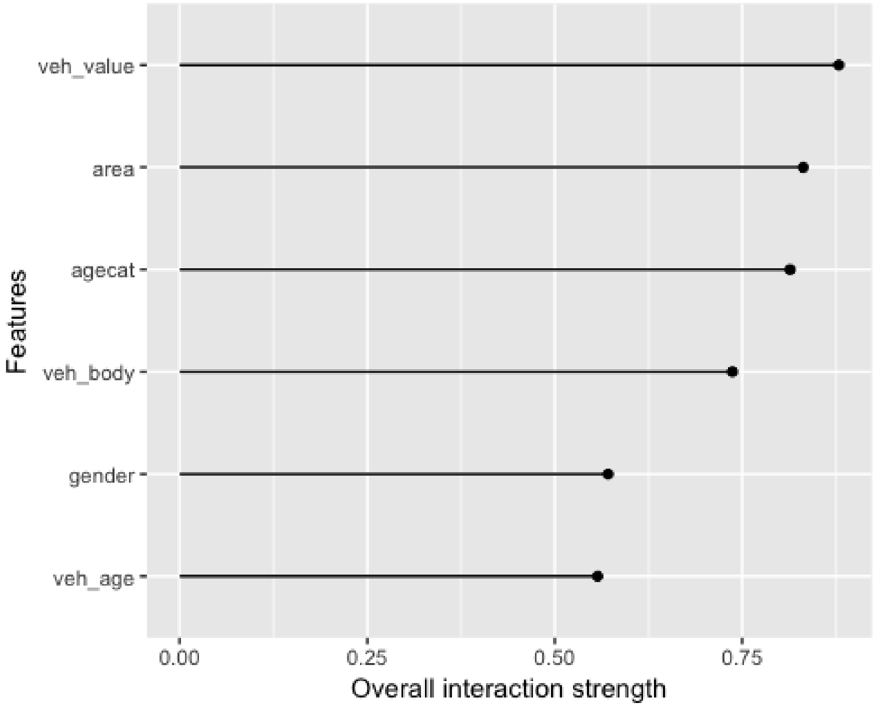
Institute and Faculty of Actuaries

Interaction Effects

- For a feature f, algorithm computes partial function only dependent on f and partial function solely dependent on each of the other features
- If variance of full (true) function can be fully explained by the sum of the above partials, no interaction is attributed to f



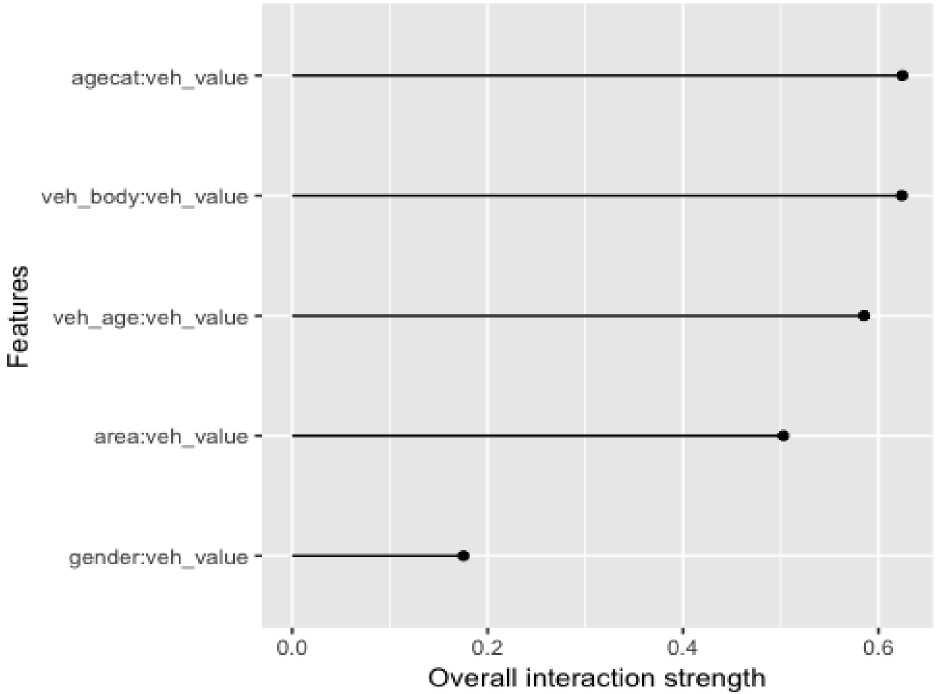
Interaction Effects – dataCar



Vehicle Value, Driver Age and Geographical Location seem to have the highest average overall interaction effects; Vehicle Body also strong

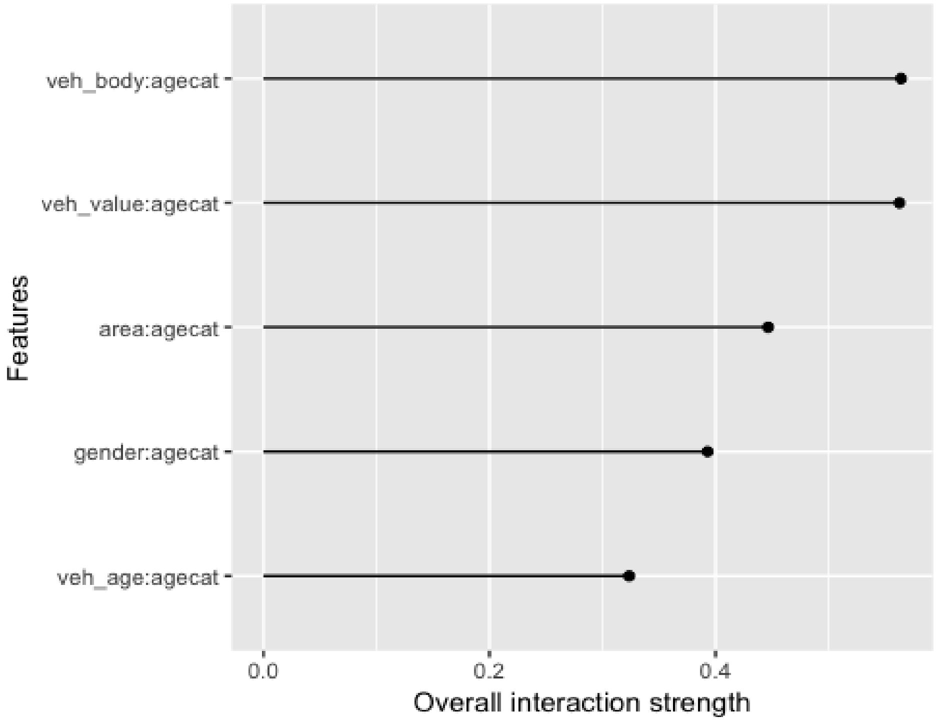
Institute and Faculty of Actuaries

Interaction Effects – dataCar



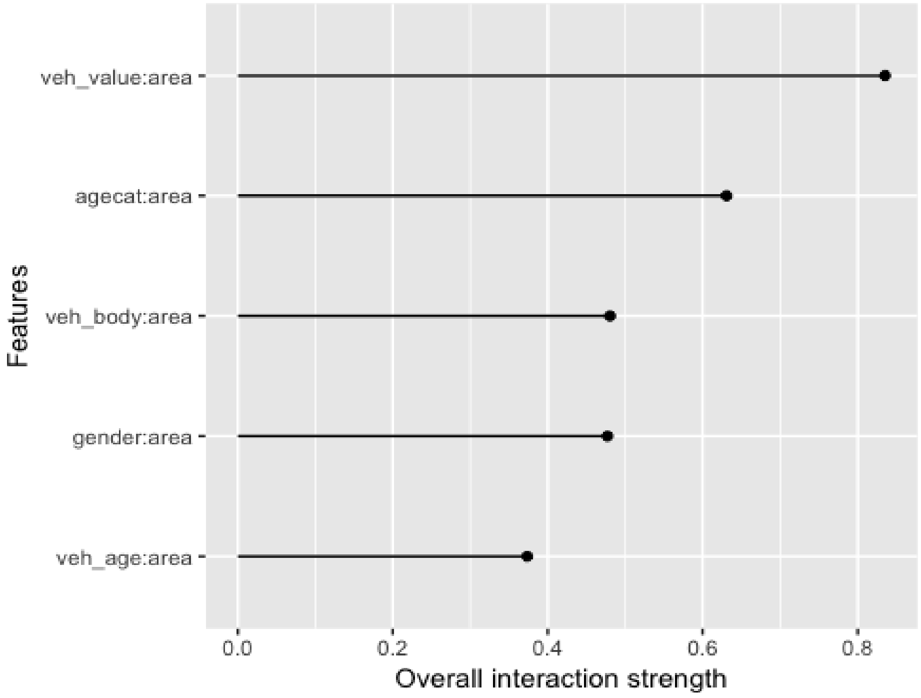
Vehicle Value – Interaction Effects

Interaction Effects – dataCar



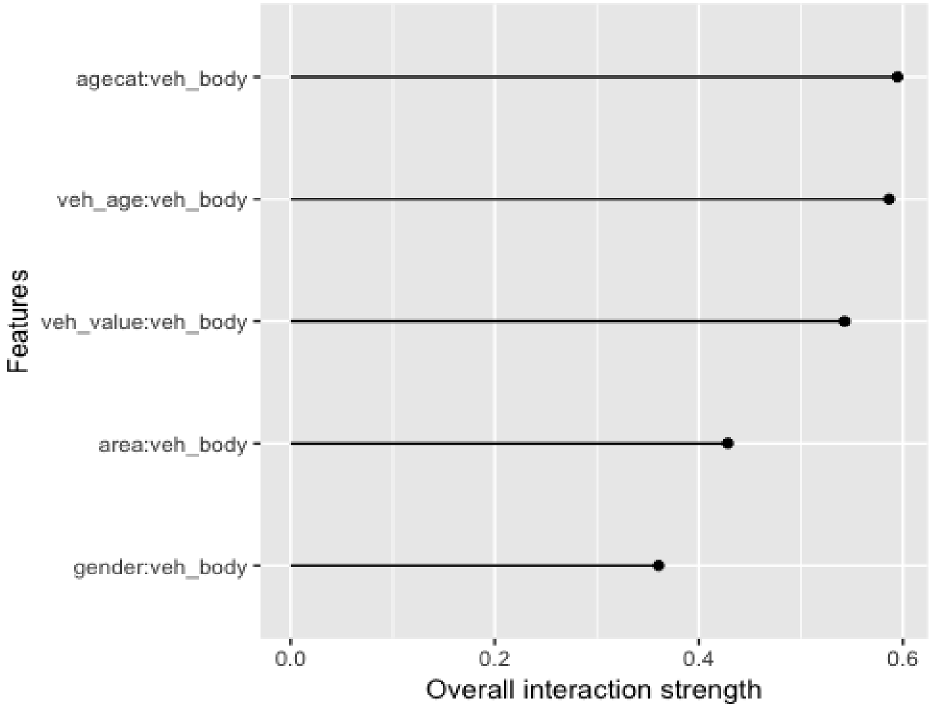
Driver Age Band – Interaction Effects

Interaction Effects – dataCar



Area – Interaction Effects

Interaction Effects – dataCar



Vehicle Body Type –
Interaction Effects

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Key Takeaways & Conclusions

18 April 2019

Agenda

Machine Learning – The Concept

Extreme Gradient Boosting

- Decision Trees
- How Gradient Boosting works

Artificial Neural Networks

- Structure and Architecture
- How ANN's Work and Learn

What does it mean to "learn"?

- Gradient Descent

Applications to Insurance Data

Interpreting Machine Learning Models

- Measuring Feature Importance
- Finding Variable Interactions

Key Takeaways and Conclusions

Institute and Faculty of Actuaries

ML: The Good and the Not-so-good

ML: The Good and the Not-so-good

- The Good:
  - Allows for complete model automation
  - No need to assume anything about the data, both in terms of rating factors and claim distributions
  - Can help us draw conclusions about hidden patterns interactions between variables

Institute and Faculty of Actuaries

ML: The Good and the Not-so-good

- The Good:
  - Allows for complete model automation
  - No need to assume anything about the data, both in terms of rating factors and claim distributions
  - Can help us draw conclusions about hidden patterns and interactions between variables
- The Not-so-good:
  - Computationally intensive – requires hardware such as GPU's and fast/powerful processors to run efficiently
  - Interpretability – Techniques are being developed to improve this

Institute and Faculty of Actuaries

Conclusions

Conclusions

- Machine Learning and AI are powerful tools which can aid actuaries in decision-making
- AI should definitely be explored and experimented with in addition to using more traditional methods such as GLM’s

Conclusions

- Machine Learning and AI are powerful tools which can aid actuaries in decision-making
- AI should definitely be explored and experimented with in addition to using more traditional methods such as GLM’s
- No one “right” model – best predictions can come from ensemble models

Institute and Faculty of Actuaries

Conclusions

- Machine Learning and AI are powerful tools which can aid actuaries in decision-making
- AI should definitely be explored and experimented with in addition to using more traditional methods such as GLM's
- No one "right" model – best predictions can come from ensemble models
- Further research being done to improve interpretability of AI, applications of Machine Learning in the actuarial realm (fraud detection, reserving)

Institute and Faculty of Actuaries

Questions

Comments

Institute and Faculty of Actuaries