アクチュアリーとデータサイエンティストが学び合うことでお互いを高め合おう

Xavier Conort
Chief Data Scientist @DataRobot
September 2018

1991-1998:

ENSAE ParisTech

université PARIS DIDEROT PARIS 7

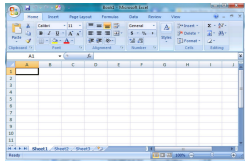
INSTITUT DES ACTUAIRES

sas

INSEE

1999-2007:

CNP ASSURANCES



Hyperion

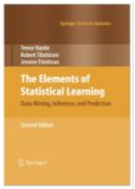
2008-2011:

CERA
Certified Enterprise Risk Actuary

AXA

BASIC RATEMAKING

2011-2013:



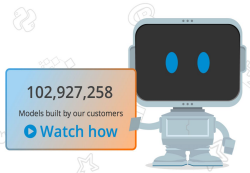
R

coursera
education for everyone

kaggle

A*STAR I2R

2013-現在:



GitHub

scikit learn

dmlc XGBoost

TensorFlow

fastTEXT

アジェンダ

- GLMとは？アクチュアリーがGLMを好んで使用するのはなぜ？
- データサイエンティストはなぜ機械学習を好む？彼らがGLMから学び損ねているものは？
- GLMの特徴をうまく組み込んだ機械学習の取り組み
- アクチュアリーが機械学習で成功するには
- 正規化された一般化線形モデルをさらに改善するには

GLMとは？アクチュアリーがGLMを好んで使用するのはなぜ？

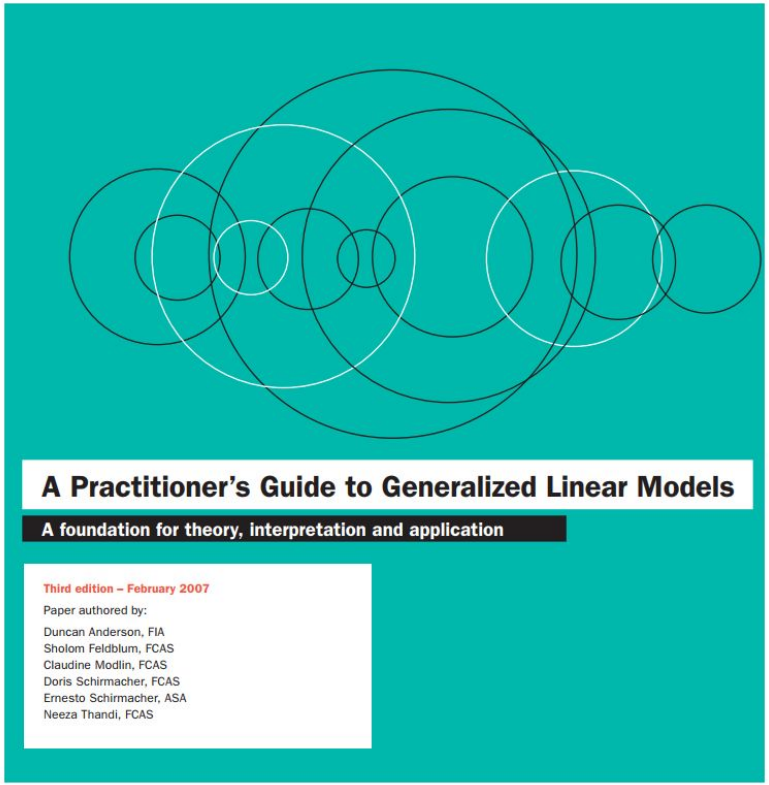
アクチュアリーが必要とするもの

アクチュアリーが対処しなければならないことは

- 強度 (severity) が**偏りのある分布**をとるリスク
- レイティングファクターと**乗法的**に変動するリスク
- **プライシングの制約**:
  - ポリシーの期間に比例した価格設定
  - 商業割引
  - 小さな変更 vs. 以前の価格設定
- **規制上の制約**:
  - 透明性
  - リスクとリスク要因の間の既知の関係
    - 例：予測されるリスクは保険料が上がると増加する

GLMの構造は、アクチュアリーのニーズに非常に関連している

GLMとは



In summary, the assumed structure of a GLM can be specified as:

$$\mu_i = E[Y_i] = g^{-1}(\sum_j X_{ij}\beta_j + \xi_i)$$

$$Var[Y_i] = \frac{\phi V(\mu_i)}{\omega_i}$$

where

Y_i is the vector of responses

$g(x)$ is the link function: a specified (invertible) function which relates the expected response to the linear combination of observed factors

X_{ij} is a matrix (the "design matrix") produced from the factors

β_j is a vector of model parameters, which is to be estimated

ξ_i is a vector of known effects or "offsets"

ϕ is a parameter to scale the function $V(x)$

$V(x)$ is the variance function

ω_i is the prior weight that assigns a credibility or weight to each observation

GLMの分散関数

$$Var(Y_i) = \frac{\phi . V(\mu_i)}{\omega_i}$$

	$V(x)$
Normal	1
Poisson	x
Gamma	x^2

アクチュアリーは**偏りのある分布**に従うリスクの強度に対処するため、ポアソン（頻度モデリング）、ガンマ（強度モデリング）またはツイーディ（コストモデリング）分布を伴うGLMを使用

ガンマ分布を使用することにより、各観測値の期待値の2乗で期待分散が増加することをモデルに通知

これが重要な理由は:

- まず、現実を反映、リスクが大きいほど分散が大きく
- また、大きな値の観測値の過学習を防ぐ
  - 実際、この期待値が高い場合、観測値と期待値との偏差にはより許容度が与えられる。

GLMのリンク関数

$$\mu_i = E[Y_i] = g^{-1}(\sum_j X_{ij}\beta_j + \xi_i)$$

ログリンク関数もアクチュアリーが実際によく使用する。そうすることで:

- 予測値が負にならないようにできる
- ほとんどの保険リスクがレイティングファクターによって**乗法的**に変動するという事実を考慮したモデルを構築できる

$$\mu_i = g^{-1}(\beta_1 x_{i1} + \ldots + \beta_p x_{ip}) = \exp(\beta_1 x_{i1}) \cdot \exp(\beta_2 x_{i2}) \ldots \exp(\beta_p x_{ip})$$

重要な注意点：ログリンク関数を備えたGLMの代わりに、Y値のログに線形モデルを適用することは、バイアスされた予測値を導くため推奨しません。log(E(Y)) は E(log(Y)) と異なります。これを行う場合は、予測を調整することを忘れないでください。そうしないと、コストの平均でなく中央値を予測してしまいます。

オフセット

オフセットは、アクチュアリーが制約を取り入れたり、モデリング戦略を強化できる素晴らしい機能:

- 予測値がエクスポージャーに確実に比例する

$$E[Y_i] = g^{-1}\left(\sum_i X_{ij}\beta_j + \xi_i\right) = \exp\left(\sum_i X_{ij}\beta_j + \ln(e_i)\right) = \exp\left(\sum_j X_{ij}\beta_j\right).e_i$$

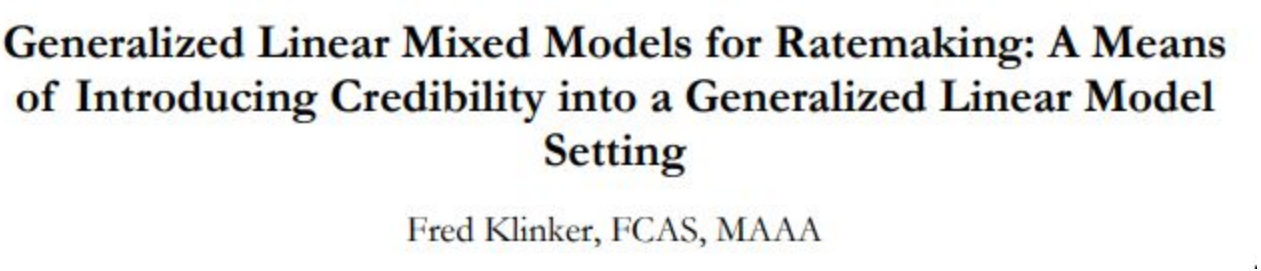
where e_i = the exposure for observation i.

- ポピュレーションの一部に任意の割引を適用できる
- 先験的な効果を取り入れ可能:
  - 他の、より大きな、似ているプロダクト
  - 市場の習慣
  - 以前の価格
- 複数の段階でモデル構築が可能:
  - 第1段階では、完全に信頼できる主要な特徴量に焦点
  - 第2段階では、信頼性がそれほど高くなく、あまり利用されない特徴量の限界効果を捉えます

その他のGLM機能

アクチュアリーが使用するもの:

- 非線形性を学習するためのスプライン
  - 手作業で定義する必要があり、計算コストがかかる
- 非線形性を捉えるためのビニング
  - ビンは統計的に十分な材料を含むように小さすぎるべきではない
- 複雑な関係性を学習するためのインタラクション
- カーディナリティの高いカテゴリ型の特徴量を処理するための混合モデル



Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting

Fred Klinker, FCAS, MAAA

データサイエンティストはなぜ機械学習を好む？彼らがGLMから学び損ねているものは？

データサイエンティストが扱う特徴量の数はGLMには多すぎる

GLMの構築には時間がかかる。アクチュアリーは、p値から、各特徴量の統計的有意性を手動でチェックし、**望ましくない特徴量をモデルから除去する**必要がある。これは、多数の特徴量やテキストなどの非構造化データが存在する場合には実用的ではない。

データサイエンティストは、機械学習アルゴリズムに組み込まれた自動正則化を好む。最も人気のある機械学習のアルゴリズムの一つは Regularized GLMs (GLMに非常に近いいとこ！) で、これはGLMの損失関数にペナルティが加えられたもの。これにより、望ましくない特徴量の係数は自動的に0に縮小される。

Regularized loss function = GLM loss function + $\lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1$

lambda2が0の場合LASSOペナルティ、lambda1が0の場合RIDGEペナルティ、どちらも0でない場合elastic-netペナルティ

データサイエンティストは複雑なシグナルを自動で学習したい

データサイエンティストは、非線形性や特徴量間の相互作用を自動的に捉えることができる機械学習アルゴリズムが大好き。

機械学習のおかげで、彼らは非常に生産的であり、多くの労力を要することなく、高い精度を得ることが可能。複雑なGLMを構築するためには、アクチュアリーはデータの知識、業務の専門知識、また大きな労力も必要。

機械学習でXヶ月からX週間にプロジェクトがスピードアップしたと聞くのは一般的。自動機械学習（自動化された前処理、自動化されたハイパーパラメータのチューニング、自動化されたモデル選択）により、これはさらにX日もしくはより短く短縮することができます。

複雑さを捉えるのに優れた人気の機械学習アルゴリズム:

- Gradient Boosting Machine
- Random Forest
- Neural Network
- Support Vector Machine

データサイエンティストが知っておくべきGLMの特徴

よく知られている

あまり知られていない

$$\mu_i = E[Y_i] = g^{-1}\left(\sum_j X_{ij}\beta_j + \xi_i\right)$$

$$Var[Y_i] = \frac{\phi V(\mu_i)}{\omega_i}$$

GLMをツールの一つとして取り入れないため、データサイエンティストは、GLMが提供できる利点を学ぶことができません:

- **大きな値を過学習するリスクが少ない** ポアソン、ガンマ、ツイーディ損失関数の使用により
- **乗法構造** ログリンク関数でアルゴリズムがより簡単にシグナルを発見
- オフセットの使用により**既知の効果、コントロールバイアス、ブースティング**を取り入れることが可能

オフセットによるバイアスのコントロールの例

GE Flight Questという米国での飛行遅延を予測するコンペに勝つために、オフセットの使用が不可欠でした。

機械にしてほしいこと:

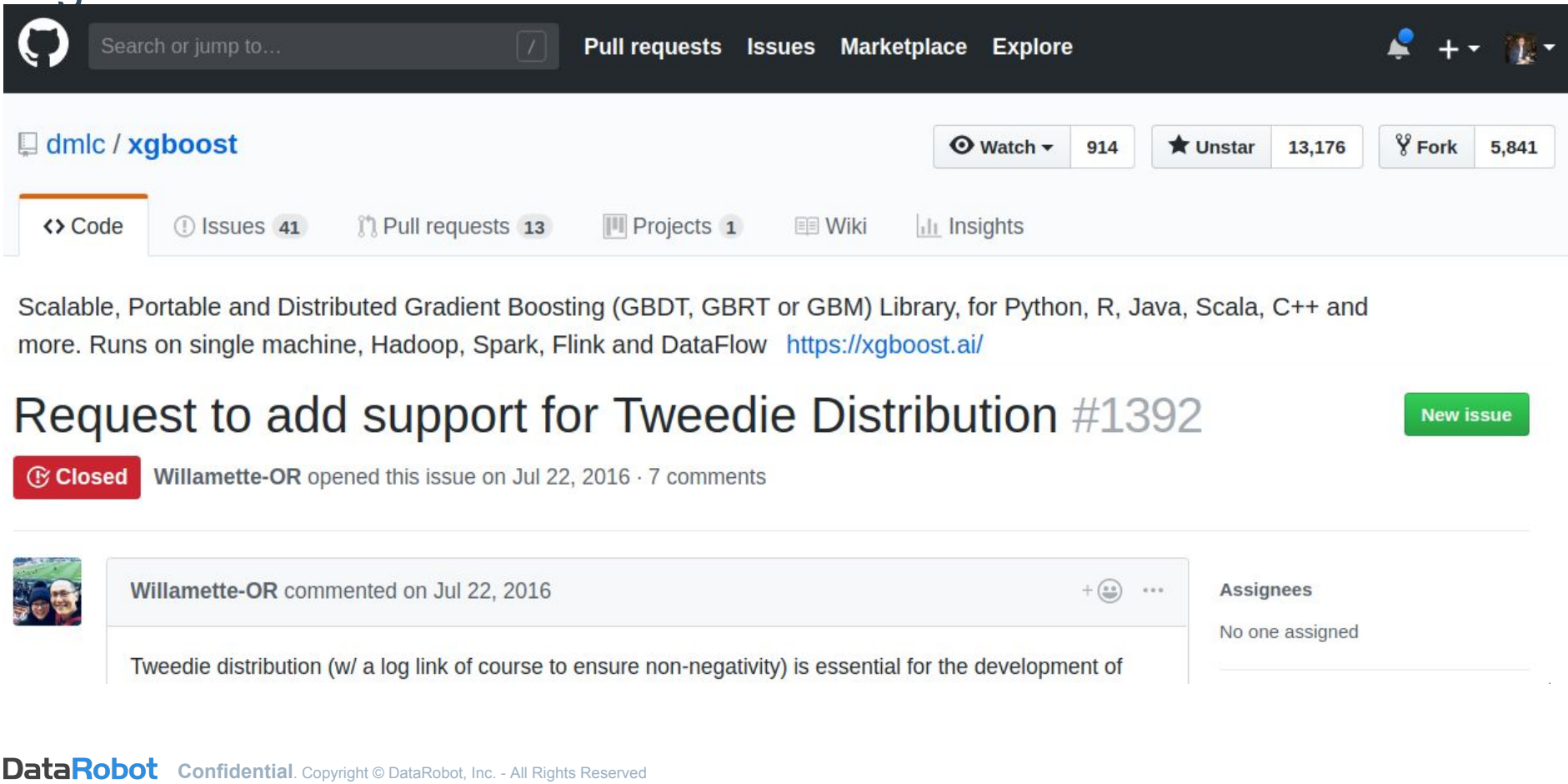
- トラフィックの悪化や悪天候のため飛行機が遅れると学習する
- ある空港が、過去3ヶ月間に一度も遅れが生じていないため、今後遅れることはないと学習しない

達成するために、2段階モデリングを使用 (アクチュアリーから学んだ方法！):

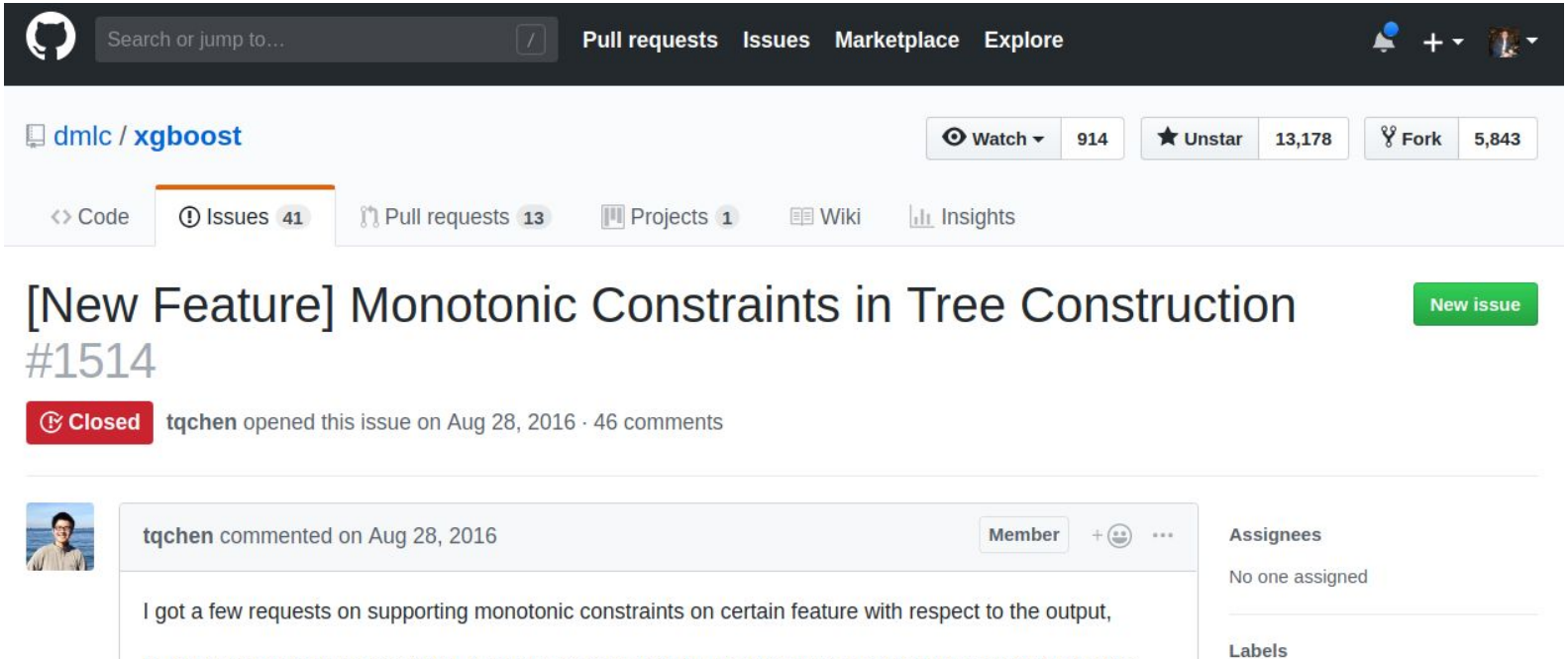
- 最初に、遅延に強く関係していそうな特徴量をGBMでフィッティング
- 次に、GBMの予測値をオフセットとして、空港名などを特徴量として、Regularized GLMでフィッティング

GLMの特徴をうまく組み込んだ機械学習の取り組み

Xgboost



Xgboost



アクチュアリーとデータサイエンティストが興奮すべき理由！

$$\mu_i = E[Y_i] = g^{-1}(\sum_j X_{ij}\beta_j + \xi_i)$$

勾配ブースティング (xgboostで実装のGradient Boosting Machine) は、
GLMに近い従兄弟です
違いは:

- デザイン行列Xはユーザによって定義されるのではなく、機械が段階的に見出す数千のルールの集合
- 係数betaは機会がゆっくり学習
- 精度の向上があまりにも低い場合にはアーリーストッピングが適用

アクチュアリーは、GLMを使用した場合と非常によく似た方法で、リスクモデリングにGBMを適用可能
一方、データサイエンティストは、彼らのツールボックスにアクチュアリーのトリックを追加可能 (指数分布、リンク関数、オフセットなど)

その他の機械学習の取り組み

オフセットのサポート
Xgboost (GBM), H20 (GBM, ElasticNet, NN), DataRobot (GBM, ElasticNet, SVM), R gbm, R glmnet (ElasticNet)

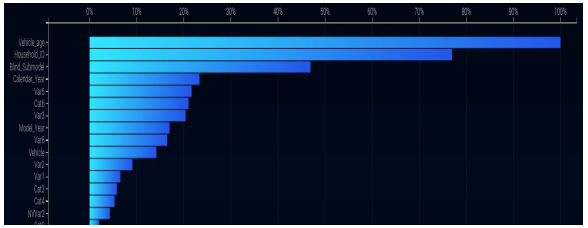
ポアソンのサポート for クレーム頻度
Xgboost (GBM), LightGBM (GBM), H20 (GBM, ElasticNet, NN), DataRobot (GBM, ElasticNet, NN, SVM), R glmnet (ElasticNet), pyglmnet (ElasticNet)

ガンマのサポート for クレーム強度
Xgboost (GBM), LightGBM (GBM), H20 (GBM, ElasticNet, NN), DataRobot (GBM, ElasticNet, NN, SVM), pyglmnet (ElasticNet)

ツイーディのサポート for クレームコスト全体 (頻度 x 強度)
Xgboost (GBM), LightGBM (GBM), H20 (GBM, ElasticNet, NN), DataRobot (GBM, ElasticNet, NN, SVM),

アクチュアリーが機械学習で成功するには

アクチュアリーへの機械学習のメリット



特徴量のインパクトをみて
新しいデータをひらめく

機械学習のおかげで、アクチュアリーは

- より生産的に => より多くのユースケース
- 既存モデルのベンチマーク
- 新しいデータの探索の高速化による
  既存のソリューションをエンリッチ
- 機械学習のインサイトにより既存のモデルの構造を改善

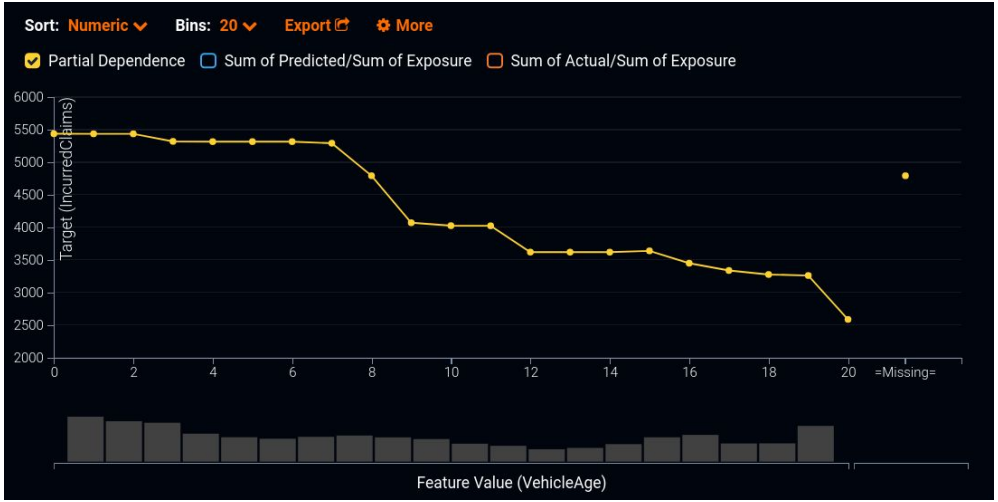
機械学習を使用してどのように既存モデルを改善できるか?

- 非線形性を捕捉するためにビニングするときに、(部分依存性プロットから)最適な
  境界を学ぶ
- 機械学習によって発見された最も影響力のある相互作用を追加
- カテゴリ特徴量のモデリングを改善

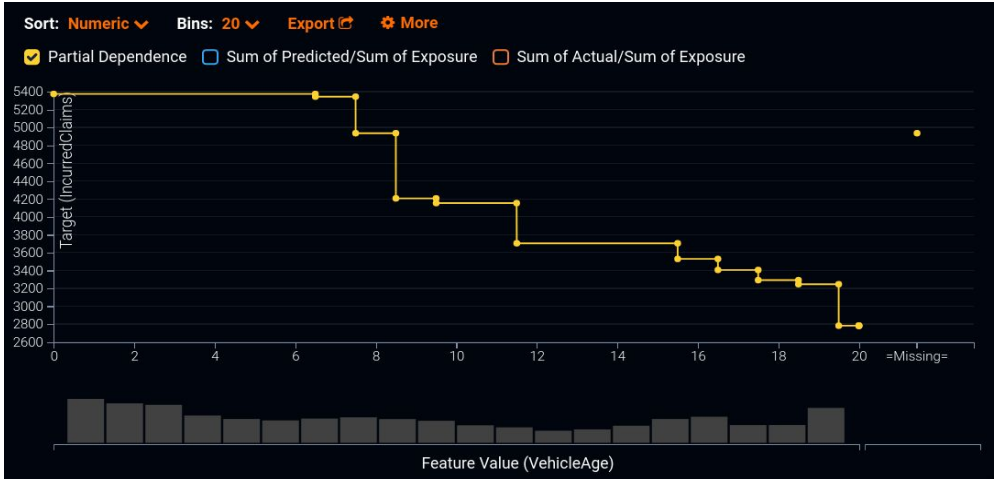
課題:GLM構造を複雑にすると、過学習のリスクが高くなる

自動でビニング

xgboostの部分依存
プロットでみる

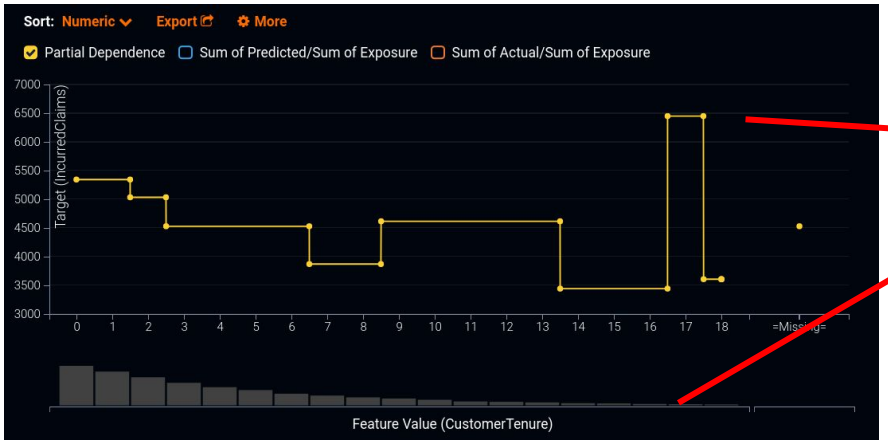


xgboostの部分依存を
近似する境界を見つける



複雑な構造を過学習するリスク

数値の特徴量を細かくビン化し、カテゴリ特徴量のレベルを追加し、インタラクションを追加すると、過学習のリスクが高くなる
残念なことに、従来の（リッジまたはラッソ）GLMを使用すると、ビンの順序の性質に関する情報が失われるため、役立たない。統計情報の少ない小さなビンの場合、リッジまたはラッソのペナルティは、隣接ビンの係数に近い値を割り当てることができない。



このビンの予測は非常にノイジー。
データが少なくて、ビンの結果を信頼できない。

複雑な構造の過学習のリスクを減らす方法

ここでは、小さなビンの過学習リスクを減らすために私がDataRobotの一般化加法モデルのために開発したソリューションを紹介します

下の方法でGLMをフィッティングするのではなく

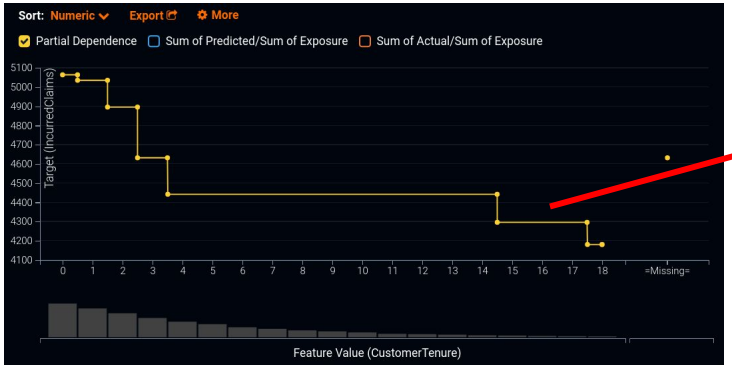
- Target ~ complex model structure, offset=O, link_function=L, distribution=d

過学習と戦うための戦略としてサロゲートモデルを使用

まず、同じターゲット、オフセット、リンク関数と分布を持つGBMに適合

次に線形モデルをフィット

- GBM_margin_predictions ~ complex model structure



結果として、ノイジーでなくなった。
GBMが正則化と特徴量の順序の性質を利用することで、データのノイズを取り込まなかった。

独自のサロゲートモデルを構築するには

- お気に入りの機械学習のモデルをフィットする:
  - 必要であればリンク関数、指数分布、オフセットをサポートするものを使用
- インサイトからフィーチャーエンジニアリングの可能性を学ぶ:
  - 特徴量のインパクトで特徴量の絞り込み
  - 各特徴量の部分依存プロットで数値型特徴量のビン化か数式を決定
- 「メインエフェクト」モデルをフィット
  - リンク関数を機械学習アルゴリズムのサンプル内の予測（マージン予測）に適用し、これをターゲットとして使用
  - 機械学習のインサイトから得た特徴量を使用して線形モデルをフィット
  - 線形モデルの予測に逆リンク関数を適用
- インタラクションを見つける:
  - 機械学習アルゴリズムによって報告されたインタラクションが存在する場合（ DataRobotでサポート）、線形モデルの予測に使用
  - もしくは「メインエフェクト」の解とあなたのお気に入りの機械学習の予測との間の残差を最もよく説明するインタラクションを探す
- 選択されたインタラクションで残差をフィットする

正規化された一般化線形モデルをさらに改善するには

ホットトピック in 保険

昨年の日本訪問の際、日本の有名なアクチュアリーの岩沢先生とお話しし、Regularized Generalized Linear Modelsのポテンシャルはまだ十分に活用されておらず、Fused Lassoはアクチュアリーにとって興味深いものだと、確信。

彼のおかげで、Fused Lassoは、GLMフレームワーク内のデータドリブンのリスク要因のビニング、レベルグルーピング、空間（または相互作用）モデリングを可能にし、小さなビンの過学習のリスクと戦うことができることを発見！

同時に、フランスにいる私の兄が、同じようなことを言っているベルギーの保険数理研究者からの非常に良いスライドを共有
https://rininsurance17.sciencesconf.org/data/Devriendt.pdf

この研究者によると、マジックは新しい罰則に由来

- Ordinal risk factors (e.g. age): Fused Lasso

$$\lambda \sum_i w_i |\beta_{i+1} - \beta_i|.$$

- Nominal risk factors (e.g. car brand and model): Generalized Fused Lasso

$$\lambda \sum_{i>k} w_{i,k} |\beta_i - \beta_k|.$$

- Spatial risk factors (e.g. postal code): Graph Guided Fused Lasso

$$\lambda \sum_{(i,k)\in G} w_{i,k} |\beta_i - \beta_k|.$$

残念ながら、ガウス分布のみをサポートするR genlassoパッケージを除き、実装はまだありません

覚えておきたいポイント

まとめ

データサイエンスは、アクチュアリーのプラクティスを受け入れるのに時間がかかりました。一方、アクチュアリーは機械学習のイノベーションに抵抗してきました。

今、両側の関心をみることができ、機械学習アルゴリズムは、アクチュアリに不可欠な機能をサポートするだけでなく、データサインティストにとっても役立つ機能をサポートします。

今後もより多くのイノベーションが期待されています。
サロゲートモデルやFused Lassoなどの新アプローチは、アクチュアリーとデータサイエンティストの両方の仕事を変える可能性のある新技術の良い例です。

質問