Naive Bayes Classifier

Explainable Machine Learning – Fabian Keck, Eva Fiserova, Natalie Martin

Wissenschaftliche Kriterien von Explainability – *Was macht ein Modell transparent?*

Simulierbarkeit

Das Bayes'sche Modell ist für die Menschen leicht zu verstehen und zu kopieren.
Bei kleineren Daten können Menschen bedingte Wahrscheinlichkeiten von Klassen und dessen Ausprägungen errechnen und Vorhersagen treffen.

Zerlegbarkeit

Die Aufschlüsselung des Modells in verständliche Teile, wie Eingaben und Berechnungen, hilft bei der Erklärung seiner Funktionsweise.
Der Datensatz kann in seine einzelnen Merkmale zerlegt werden. Wahrscheinlichkeiten können für Merkmale und deren Kookkurrenzen berechnet werden.

Algorithmische Transparenz

Das Modell zeigt uns genau, wie es Entscheidungen trifft, so dass der Prozess nachvollziehbar ist.
Die fit()-Funktion von Scitkitlearn verhindert Einblicke in den Trainingsprozess, allerdings können Wahrscheinlichkeiten abgefragt werden.

Vorgehen & Accuracy

Es wurden Methoden angewendet, um die Accuracy zu verbessern.

Nur Gaussian Naive	80%
Reduziert auf die wichtigsten Merkmale	81%
Mit Hyperparametertuning	83%
Mit PCA	82%

Der Datensatz wurde auf die wichtigsten Merkmale reduziert. Dabei wurden verschiedene Anzahlen von Merkmalen getestet.

Bayes Theorem & Bayes Classifier

Das Bayes Theorem basiert auf dem Konzept der bedingten Wahrscheinlichkeiten.

Posterior **Likelihood** **Prior Class**

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Prior Feature

Der naive Bayes Classifier verwendet dieses Theorem um im Trainingsprozess die Wahrscheinlichkeit für die Zugehörigkeit zu einer Klasse A für eines bestimmtes Merkmales B zu bestimmen. Eine Prämisse hierfür ist die Unabhängigkeit der Merkmale.

Das Theorem im Trainingsprozess:
1. Wahrscheinlichkeit der Klasse P(A)* – **Prior Class**
2. Umgekehrte/Inverse bedingte Wahrscheinlichkeit P(B|A)* – **Likelihood**
3. Wahrscheinlichkeit des Merkmales P(B)* – **Prior Feature**
4. Berechnung der gesuchte Wahrscheinlichkeit P(A|B) – **Posterior**

*anfängliche Wahrscheinlichkeit (Historische Daten, Annahmen)

Explainable Boosting Machine (EBM)

Explainable Boosting Machine (EBM) ist ein baumbasiertes, zyklisches Gradient Boosting Generalized Additive Model. EBM ist ein Glassbox-Modell mit hoher Genauigkeit, das mit den modernsten Methoden des maschinellen Lernens wie Random Forest vergleichbar ist, und gleichzeitig hochgradig intelligent und erklärbar ist.

Als erklärbares Modell wird dieses zum Vergleich hinzugezogen.

Bayes vs. EBM

$P(A\|B) = \frac{P(B\|A) \cdot P(A)}{P(B)}$	$g(E[y]) = \beta_0 + \sum f_i(x_i) + \sum f_{i,j}(x_i, x_j)$
Accuracy bei Dataset: 80%	Accuracy bei Dataset: 87%
Verteilung der Daten kann mit sns.countplot analysiert werden, für jede einzelne Variable	Funktion Class Histogram: Analyse der Verteilung der Daten
Marginal Plot mit seaborn Funktion und Berechnung des Pearson-Korrelationskoeffizienten, für jede einzelne Variable	Funktion Marginal: Verteilung der Datenmerkmale, Berechnung des Pearson-Korrelationskoeffizienten

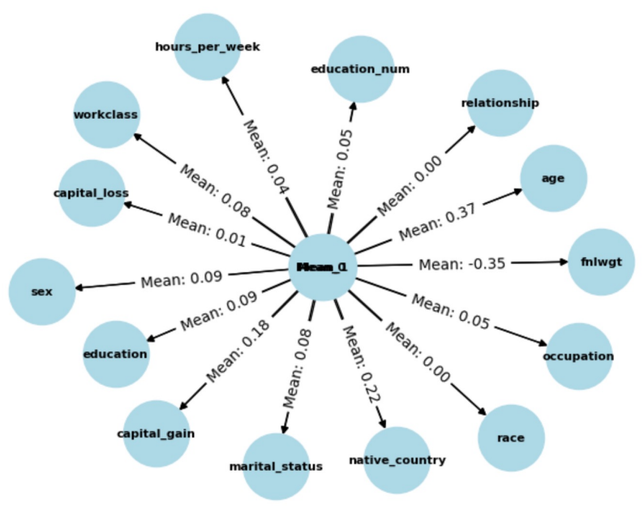
Bayes vs. EBM

Feature Importance kann mit plt gezeigt werden, Beitrag, den jedes Merkmal zu den Vorhersagen leistet, nicht	Explain.global(): zeigt Feature Importance + Beitrag, den jedes Merkmal zu den Vorhersagen leistet
Keine ähnliche Funktion	Explain.local(): zeigt zeigt die Aufschlüsselung, wie viel jeder Begriff zur Vorhersage für eine einzelne Stichprobe beiträgt

Fazit: Da EBM speziell auf die Interpretierbarkeit ausgelegt ist, liefert es viele Funktionen, mit welchen man sowohl das Dataset als auch die Ergebnisse des trainierten Modells betrachten kann. Auch beim Gaussian Naive Bayes kann man viele Funktionen anwenden, die die Explainability unterstützen, jedoch werden diese Funktionen von anderen Bibliotheken bereitgestellt und man muss sie oft für jede Variable einzeln ausführen.

Explainability & Dataset

Der Directed Acyclic Graphs (DAG) zeigt die mittleren Werte der Merkmale in Bezug auf das Einkommen und bietet verständliche Einblicke in die Einflussfaktoren auf das Zielattribut. Diese Graphik unterstützt das erklärbare Maschinenlernen, indem sie wichtige Merkmale hervorhebt und deren Auswirkungen auf die Einkommensvorhersage veranschaulicht



Fazit

Der Naive Bayes Classifier ist für seine Einfachheit und leichte Implementierung bekannt, bietet jedoch nicht dasselbe Maß an Interpretierbarkeit wie einige andere Modelle, insbesondere solche, die ausdrücklich auf Interpretierbarkeit ausgelegt sind.

Naive Bayes Classifier

Explainable Machine Learning – Fabian Keck, Eva Fiserova, Natalie Martin

Wissenschaftliche Kriterien von Explainability – *Was macht ein Modell transparent?*

Simulierbarkeit

Das Bayes'sche Modell ist für die Menschen leicht zu verstehen und zu kopieren.
Bei kleineren Daten können Menschen bedingte Wahrscheinlichkeiten von Klassen und dessen Ausprägungen errechnen und Vorhersagen treffen.

Zerlegbarkeit

Die Aufschlüsselung des Modells in verständliche Teile, wie Eingaben und Berechnungen, hilft bei der Erklärung seiner Funktionsweise.
Der Datensatz kann in seine einzelnen Merkmale zerlegt werden. Wahrscheinlichkeiten können für Merkmale und deren Kookkurrenzen berechnet werden.

Algorithmische Transparenz

Das Modell zeigt uns genau, wie es Entscheidungen trifft, so dass der Prozess nachvollziehbar ist.
Die fit()-Funktion von Scitkitlearn verhindert Einblicke in den Trainingsprozess, allerdings können Wahrscheinlichkeiten abgefragt werden.

Vorgehen & Accuracy

Es wurden Methoden angewendet, um die Accuracy zu verbessern.

Nur Gaussian Naive	80%
Reduziert auf die wichtigsten Merkmale	81%
Mit Hyperparametertuning	83%
Mit PCA	82%

Der Datensatz wurde auf die wichtigsten Merkmale reduziert. Dabei wurden verschiedene Anzahlen von Merkmalen getestet.

Bayes Theorem & Trainingsprozess

Das Bayes Theorem basiert auf dem Konzept der bedingten Wahrscheinlichkeiten.

Posterior **Likelihood** **Prior Class**

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Prior Feature

Der Trainingsprozess beinhaltet das Errechnen der Wahrscheinlichkeiten. Das geschieht auch unter Verwendung des Theorems. In Bezug auf den Classifier werden die Klassen A und die Merkmale B betrachtet. **Das Ziel ist es, P(A|B) zu schätzen, also die Wahrscheinlichkeit einer bestimmten Klasse, wenn bestimmte Merkmale zu beobachten sind.**

Das Theorem im Trainingsprozess:
1. Schätzung der Klassenwahrscheinlichkeiten P(A) – **Prior Class**
2. Schätzung der klassebedingten Wahrscheinlichkeiten P(B|A) – **Likelihood**
3. Anfängliche Wahrscheinlichkeit oder Annahme über das Eintreten eines Ereignisses – **Prior Feature**
4. Anwendung des Bayes-Theorems P(A|B) – **Posterior**

Explainable Boosting Machine (EBM)

Explainable Boosting Machine (EBM) ist ein baumbasiertes, zyklisches Gradient Boosting Generalized Additive Model. EBM ist ein Glassbox-Modell mit hoher Genauigkeit, das mit den modernsten Methoden des maschinellen Lernens wie Random Forest vergleichbar ist, und gleichzeitig hochgradig intelligent und erklärbar ist.

Als erklärbares Modell wird dieses zum Vergleich hinzugezogen.

Bayes vs. EBM

$P(A\|B) = \frac{P(B\|A) \cdot P(A)}{P(B)}$	$g(E[y]) = \beta_0 + \sum f_i(x_i) + \sum f_{i,j}(x_i, x_j)$
Accuracy bei Dataset: 80%	Accuracy bei Dataset: 87%
Verteilung der Daten kann mit sns.countplot analysiert werden, für jede einzelne Variable	Funktion Class Histogram: Analyse der Verteilung der Daten
Marginal Plot mit seaborn Funktion und Berechnung des Pearson-Korrelationskoeffizienten, für jede einzelne Variable	Funktion Marginal: Verteilung der Datenmerkmale, Berechnung des Pearson-Korrelationskoeffizienten

Bayes vs. EBM

Feature Importance kann mit plt gezeigt werden, Beitrag, den jedes Merkmal zu den Vorhersagen leistet, nicht	Explain.global(): zeigt Feature Importance + Beitrag, den jedes Merkmal zu den Vorhersagen leistet
Keine ähnliche Funktion	Explain.local(): zeigt zeigt die Aufschlüsselung, wie viel jeder Begriff zur Vorhersage für eine einzelne Stichprobe beiträgt

Fazit: Da EBM speziell auf die Interpretierbarkeit ausgelegt ist, liefert es viele Funktionen, mit welchen man sowohl das Dataset als auch die Ergebnisse des trainierten Modells betrachten kann. Auch beim Gaussian Naive Bayes kann man viele Funktionen anwenden, die die Explainability unterstützen, jedoch werden diese Funktionen von anderen Bibliotheken bereitgestellt und man muss sie oft für jede Variable einzeln ausführen.

Explainability

Der Naive Bayes Classifier ist für seine Einfachheit und leichte Implementierung bekannt, bietet jedoch nicht dasselbe Maß an Interpretierbarkeit wie einige andere Modelle, insbesondere solche, die ausdrücklich auf Interpretierbarkeit ausgelegt sind.

Fazit

Schritftgröße 15

Naive Bayes Classifier

Explainable Machine Learning – *Wintersemester 23/24*

Wissenschaftlicher Ansatz – Kriterien von Explainability

Simulierbarkeit

Das Bayes'sche Modell ist für die Menschen leicht zu verstehen und zu kopieren. *Bei kleineren Daten können Menschen bedingte Wahrscheinlichkeiten von Klassen und dessen Ausprägungen errechnen und Vorhersagen treffen.*

Zerlegbarkeit

Die Aufschlüsselung des Modells in verständliche Teile, wie Eingaben und Berechnungen, hilft bei der Erklärung seiner Funktionsweise. *Der Datensatz kann in seine einzelnen Merkmale zerlegt werden. Wahrscheinlichkeiten können für Merkmale berechnet werden.*

Algorithmische Transparenz

Das Modell zeigt uns genau, wie es Entscheidungen trifft, so dass der Prozess nachvollziehbar ist. *Die fit()-Funktion von Scitkitlearn verhindert Einblicke in den Trainingsprozess, allerdings können Wahrscheinlichkeiten abgefragt werden.*

Bayes Theorem

basiert auf bedingten Wahrscheinlichkeiten

Posterior, Likelihood, Prior

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- Posterior: wird im Nachhinein berechnet
- Prior
- Likelihood:
- XY:

Vorgehen & Accuracy

Es wurden Methoden angewendet, um die Accuracy zu verbessern.

Nur Gaussian Naive	80%
Reduziert auf die wichtigsten Merkmale	81%
Mit Hyperparametertuning	83%
Mit PCA	82%

Der Datensatz wurde auf die wichtigsten Merkmale reduziert. Dabei wurden verschiedene Anzahlen von Merkmalen getestet.

Explainability (Daten)

Schritftgröße 15

Explainable Boosting Machine

Explainable Boosting Machine (EBM) ist ein baumbasiertes, zyklisches Gradient Boosting Generalized Additive Model. EBM ist ein Glassbox-Modell mit hoher Genauigkeit, das mit den modernsten Methoden des maschinellen Lernens wie Random Forest vergleichbar ist, und gleichzeitig hochgradig intelligent und erklärbar ist.

Als erklärbares Modell wird dieses zum Vergleich hinzugezogen.

Gaussian Naive Bayes Classifier

Posterior, Prior

$$P(class|data) = \frac{P(data|class) \cdot P(class)}{P(data)}$$

Bayes vs. XBC

Schritftgröße 15

Interpretierbarkeit der Daten in der Anfangsphase der Modellierung: bei beiden Modellen gleich

Trainingsprozess

Der Trainingsprozess beinhaltet das Errechnen der Wahrscheinlichkeiten. Das geschieht auch unter Verwendung des Theorems. In Bezug auf den Classifier werden die Klassen A und die Merkmale B betrachtet. **Das Ziel ist es, P(A|B) zu schätzen, also die Wahrscheinlichkeit einer bestimmten Klasse, wenn bestimmte Merkmale zu beobachten sind.**

Der Trainingsprozess:

1. Schätzung der Wahrscheinlichkeiten P(A): Dies sind die Wahrscheinlichkeiten für jede Klasse, unabhängig von den Merkmalen.
2. Schätzung der klassebedingten Wahrscheinlichkeiten P(B|A): Es wird angenommen, dass die Merkmale in jeder Klasse normalverteilt sind. Das bedeutet, dass für jede Klasse die Mittelwerte und Standardabweichungen für jedes Merkmal errechnet werden.
3. Anwendung des Bayes-Theorems P(A|B): Mit den geschätzten Wahrscheinlichkeiten werden dann die Wahrscheinlichkeiten P(A|B) berechnet, die für die Vorhersagen verwendet werden.

Fazit

Schritftgröße 15

Nati, Evi & Fabi