

Bayesian Networks на примере Zoo Dataset

Студент: Зиновьев Максим
Группа: 307Б

Что такое Bayesian Networks

- - Вероятностная графовая модель.
- - Узлы — переменные, рёбра — вероятностные зависимости.
- - Используются для интерпретации, прогнозов, принятия решений.
- Формула Байеса:
- $P(A|B) = P(B|A) * P(A) / P(B)$

Описание набора данных Zoo

- - 101 животное, 17 бинарных признаков.
- - Примеры признаков: hair, feathers, milk, backbone, legs.
- - Цель: предсказать класс животного (1–7).

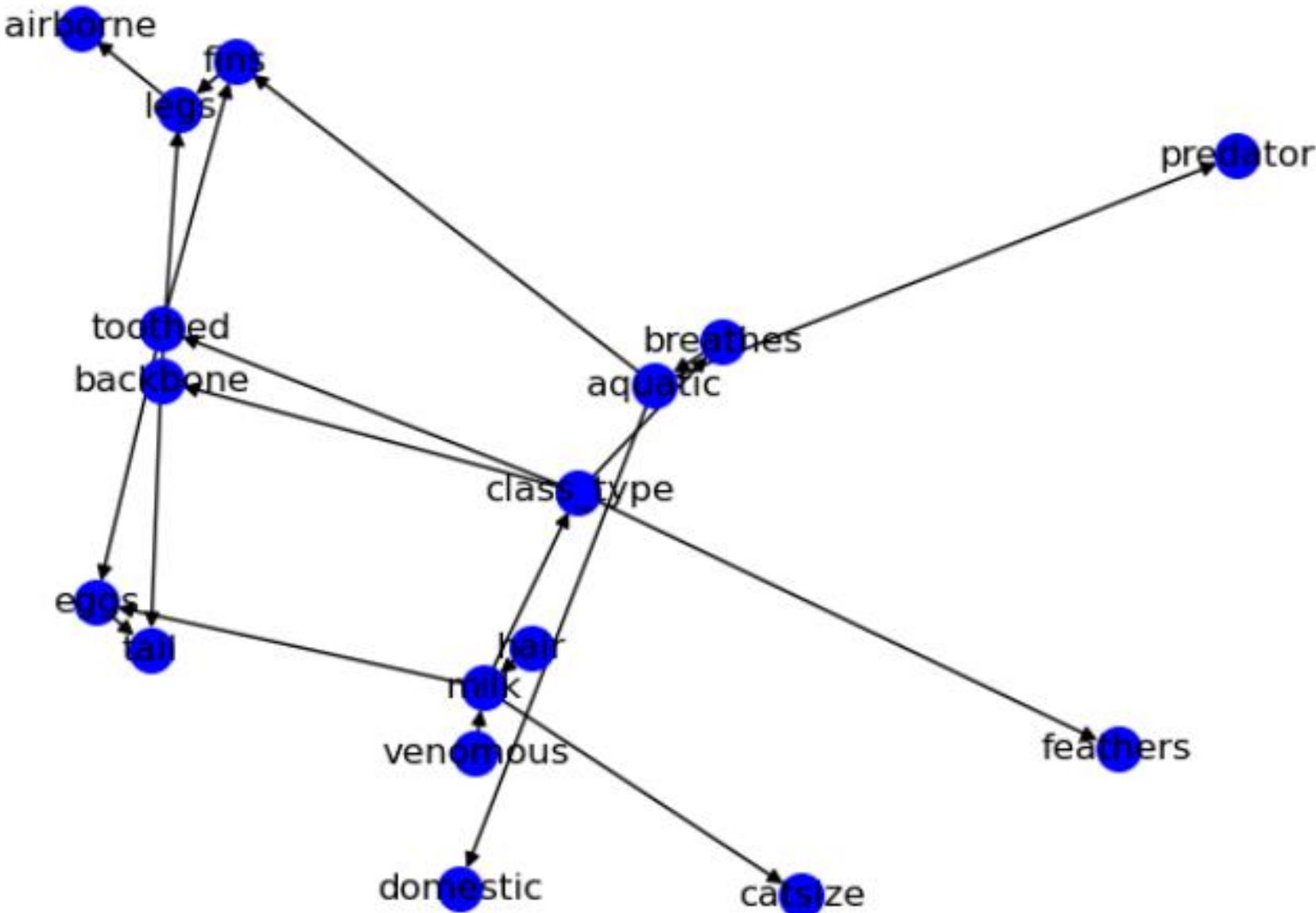
Загрузка и предварительная обработка

- Код загрузки:
- `df = pd.read_csv('zoo.csv')`
- `df = df.drop(columns=['animal_name'])`
- Все признаки уже категориальные и числовые.

Построение структуры сети

- Использован HillClimbSearch + BicScore:
- `hc = HillClimbSearch(df)`
- `model = hc.estimate(BicScore(df))`
- - Автоматически нашёл зависимые признаки.
- - Центр графа — `class_type`.

Bayesian Network for Zoo Animal Classification



Выводы

- - Байесовская сеть выявила интерпретируемые зависимости.
- - BN немного хуже baseline по метрикам.
- - Преимущество BN — объяснимость структуры.
- - Naive Bayes обеспечивает максимальную точность на Zoo.