

Ziel: Implementierung von Active Learning mit den Ansätzen Uncertainty Sampling (Exploitation) und Diversity Sampling (Exploration) und Vergleich der Ergebnisse.

Schritte:

Implementieren Sie Active Learning Ansätze Ihrer Wahl (je 2 Varianten von Exploitation und Exploration) und wenden Sie diese auf einen gewählten Datensatz an. Vergleichen Sie die Leistung der Active Learning Ansätze mit einem herkömmlichen Ansatz, bei dem das Modell auf dem gesamten Trainingsdatensatz trainiert wird. Messen Sie die Genauigkeit aller Ansätze auf einem Testdatensatz.

1. Laden Sie den gegebenen Datensatz (der für Active Learning geeignet ist) und teilen Sie ihn in Trainings- und Testdaten auf.
2. Wählen Sie ein Klassifikationsmodell Ihrer Wahl aus (z. B. Random Forest, Support Vector Machine, Neuronales Netzwerk).
3. Teilen Sie den Datensatz in Trainings- und Testdaten auf (z.B. im Verhältnis 80:20).
4. Trainieren Sie Ihren Datensatz ohne Active Learning (Accuracy merken/speichern für den späteren Vergleich)
5. Implementieren Sie die Funktionen für Uncertainty Sampling und Diversity Sampling:
 - Uncertainty Sampling: Wählen Sie die Datenpunkte aus, für die das Modell am unsichersten ist (z.B. höchste Klassifikationsunsicherheit).
 - Diversity Sampling: Wählen Sie die Datenpunkte aus, die die größte Distanz zu den bereits ausgewählten Datenpunkten haben (z.B. mittels Distanzmetriken wie Euclidean Distance).
 - [optional] Probieren Sie die Kombination von Uncertainty und Diversity Sampling aus
 - Liefert dieser Ansatz noch bessere Ergebnisse?
6. Implementieren Sie die Aktualisierung der Trainingsdaten (Re-Training) mit den ausgewählten Datenpunkten (kann nach jeder Iteration sein oder wie gewünscht z.B.: nach 5 Iterationen)
7. Führen Sie Active Learning durch:
 - Wenden Sie Uncertainty Sampling an, um eine bestimmte Anzahl von Datenpunkten auszuwählen.
 - Aktualisieren Sie die Trainingsdaten und trainieren Sie das Modell neu.
 - Evaluieren Sie die Genauigkeit des Modells auf den Testdaten.
 - Wenden Sie Diversity Sampling an, um eine bestimmte Anzahl von Datenpunkten auszuwählen.
 - Aktualisieren Sie erneut die Trainingsdaten und trainieren Sie das Modell neu.
 - Evaluieren Sie erneut die Genauigkeit des Modells auf den Testdaten.
8. Vergleichen Sie die Genauigkeiten der Ansätze (Uncertainty Sampling und Diversity Sampling).
9. Diskutieren Sie die Ergebnisse und reflektieren Sie über die Vor- und Nachteile der beiden Ansätze.

Möglicher Datensatz: MNIST, ein Datensatz von handgeschriebenen Ziffern.

Vorgegebenes Modell: Sie können ein beliebiges Klassifikationsmodell Ihrer Wahl verwenden.

Hinweise:

- Sie können vorhandene Bibliotheken wie scikit-learn verwenden, um das Klassifikationsmodell zu initialisieren und zu trainieren.
- Implementieren Sie die Funktionen für Uncertainty Sampling und Diversity Sampling selbst (Vorlage github von erwähnten Buch)
- Experimentieren Sie mit unterschiedlichen Parameterwerten (Anzahl der ausgewählten Datenpunkte, Metriken für Diversity Sampling usw.) und analysieren Sie ihre Auswirkungen auf die Genauigkeit des Modells.