**Resumen**

En este estudio, se utilizó el conjunto de datos Iris para entrenar un modelo de clasificación multiclase utilizando la técnica de Clasificación por Máquina de Vectores de Soporte con Codificación de Salida-Entrada Corregida (ECOC). Se dividió el conjunto de datos en entrenamiento (70%) y prueba (30%) y se evaluó el desempeño del modelo mediante la exactitud, la matriz de confusión, la precisión, el recall y el F1 Score. Además, se generaron curvas de aprendizaje para visualizar el error de entrenamiento y prueba a medida que aumenta el número de ejemplos de entrenamiento. Los resultados demostraron una alta exactitud y buenos valores de precisión, recall y F1 Score para cada clase.

**Palabras Clave**

Clasificación Multiclase, ECOC, Iris Dataset, Matriz de Confusión, Precisión, Recall, F1 Score, Curvas de Aprendizaje.

**Introducción**

La clasificación multiclase es un desafío común en el campo del aprendizaje automático, donde el objetivo es categorizar instancias en una de las múltiples clases posibles. En este estudio, utilizamos el conjunto de datos Iris, que es un conjunto de datos clásico en la investigación del aprendizaje automático, para entrenar un modelo de clasificación multiclase utilizando la técnica de Codificación de Salida-Entrada Corregida (ECOC) aplicada a una Máquina de Vectores de Soporte (SVM).

El conjunto de datos Iris contiene 150 muestras de flores de iris, cada una con cuatro características: longitud del sépalo, anchura del sépalo, longitud del pétalo y anchura del pétalo. Estas muestras pertenecen a una de las tres especies: Iris setosa, Iris versicolor e Iris virginica. El objetivo es predecir la especie de una flor de iris basada en las características dadas.

**Desarrollo**

Preparación del Conjunto de Datos

El conjunto de datos Iris se cargó y se dividió en características (X) y etiquetas (Y). Las etiquetas categóricas se convirtieron a valores numéricos para facilitar el proceso de clasificación. Posteriormente, se dividió el conjunto de datos en un 70% para entrenamiento y un 30% para prueba.

A black screen with white text

Description automatically generated

Entrenamiento del Modelo

Se entrenó un modelo de clasificación multiclase utilizando la técnica ECOC aplicada a una Máquina de Vectores de Soporte (SVM) con los datos de entrenamiento.

A black background with white text

Description automatically generated  
  
Evaluación del Modelo

Se evaluó el modelo utilizando los datos de prueba para predecir las etiquetas y calcular la exactitud. También se generó una matriz de confusión y se calcularon las métricas de precisión, recall y F1 Score para cada clase.

A computer screen with white text

Description automatically generated

Generación de Curvas de Aprendizaje

Se generaron curvas de aprendizaje para visualizar el error de entrenamiento y prueba a medida que aumenta el número de ejemplos de entrenamiento.

A computer screen with white text

Description automatically generated

**Análisis y Resultados**

Exactitud del Modelo

El modelo alcanzó una exactitud del 97.78% en el conjunto de datos de prueba, lo que indica un alto desempeño en la clasificación de las especies de iris.

Matriz de Confusión

La matriz de confusión mostró una distribución precisa de las predicciones entre las tres clases, con pocas confusiones entre ellas.

Métricas Adicionales

Las métricas de precisión, recall y F1 Score para cada clase fueron satisfactorias, demostrando que el modelo tiene un buen equilibrio entre precisión y capacidad de recuperación.

Curvas de Aprendizaje

Las curvas de aprendizaje mostraron que el error de entrenamiento disminuye a medida que aumenta el número de ejemplos, mientras que el error de prueba se estabiliza, indicando que el modelo generaliza bien los datos no vistos.

Visualización de Datos

Se realizaron gráficos de dispersión tridimensional y bidimensional para visualizar la distribución de las muestras en el espacio de características, mostrando una clara separación entre las diferentes clases.

A screenshot of a computer

Description automatically generated

**Referencias**

Fisher, R.A. (1936). "The use of multiple measurements in taxonomic problems". Annals of Eugenics, 7(2), 179-188.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.

MATLAB Documentation. (n.d.). fitcecoc: Train a multiclass model using error-correcting output codes. Retrieved from MathWorks.