

Proyecto 2: Análisis de Métodos de Reducción de Dimensionalidad para la Clasificación de imágenes

Prof. Cristian López Del Álamo

1. Introducción

Este proyecto se centra en la clasificación de imágenes utilizando un conjunto de datos estandarizado y ampliamente adoptado en el campo del aprendizaje automático para tareas de clasificación multiclase. Este dataset de imágenes está compuesto por 70,000 muestras en escala de grises.

Cada imagen tiene una baja resolución de 28×28 píxeles, y representa diferentes categorías de prendas de vestir. El conjunto está dividido en 10 clases distintas (por ejemplo, calzado, prendas superiores e inferiores). Al ser una imagen de 28×28 , cada muestra se traduce en un vector de 784 características (píxeles), lo que configura un espacio de alta dimensionalidad.

Trabajar con datos de tan alta dimensión puede resultar ineficiente e incluso perjudicial para el desempeño de los modelos de clasificación. Este fenómeno, en el que el espacio de características es vasto y disperso, se relaciona con la maldición de la dimensionalidad.

El objetivo principal de este proyecto es mitigar este desafío explorando y comparando diversas técnicas de reducción de dimensionalidad. Se busca transformar el espacio de 784 características en un espacio más pequeño que retenga la información esencial, para luego evaluar qué combinación de (técnica de reducción + clasificador) ofrece el mejor balance entre rendimiento y eficiencia sobre este conjunto de imágenes de moda.

2. Importancia y Justificación del Proyecto

- Optimización de modelos: Reducir la cantidad de características puede acelerar el entrenamiento, disminuir el riesgo de sobreajuste y facilitar la visualización de los datos sin comprometer la precisión.
- Aprendizaje visual e interpretabilidad: Métodos como t-SNE y UMAP permiten visualizar cómo se agrupan las clases en un espacio bidimensional, ayudando a comprender la estructura interna del conjunto de datos y su separabilidad.

3. Actividades a Desarrollar

El proyecto se desarrollará en cuatro fases principales:

1. Fase 1: Preparación del entorno y carga de datos

- Configuración del entorno en Python (Jupyter Notebook, Scikit-learn, TensorFlow).
- Descarga y carga del dataset.
- Preprocesamiento: normalización de valores de píxeles (0 a 1) y división en conjuntos de entrenamiento (80 %) y prueba (20 %).

2. Fase 2: Aplicación de reducción de dimensionalidad

- Implementación de los siguientes métodos:
 - Lineales: PCA, NMF.
 - No lineales: t-SNE, UMAP.
 - Investigar: Spectral Embedding e Isomap.
- Generación de visualizaciones bidimensionales (2D) de los datos reducidos.

3. Fase 3: Clasificación y evaluación

- Entrenamiento de clasificadores sobre cada conjunto reducido:
 - SVM
 - Regresión Logística
 - Random Forest
 - KNN
- Evaluación mediante métricas: Accuracy, Precision, Recall y F1-Score.

4. Fase 4: Análisis comparativo y conclusiones

- Creación de una tabla comparativa de resultados.
- Análisis crítico: ¿qué combinación reducción + clasificador fue más efectiva?
- Elaboración del informe final y discusión de resultados.

4. Documento

Directrices del Documento

Límite de Páginas: El documento no debe exceder las **8 hojas**.

- Enfocarse en la Metodología y los Resultados.
- Usar gráficos y tablas claras para la comparación de modelos.
- Utilize este **template** para su documento
- **El mejor trabajo será publicado en la revista GINIA de UTEC.**
- En caso de detección de copia, todo el equipo tendrá cero en la nota sin derecho a reclamo.
- Puede usar librerías, pero el desarrollo debe ser hecho por todo el equipo.
- **No olvide colocar el % de participación al lado del nombre de cada integrante. Si esto no está la nota se divide entre el número de participantes.**

5. Rúbrica de Evaluación

Criterio		Excelente (4 pts)	Bueno (3 pts)	Suficiente (2 pts)	Insuficiente (0-1 pts)
Introducción y Justificación	y	Define claramente el problema, los objetivos y la motivación del proyecto.	Presenta adecuadamente el contexto y objetivos, pero con justificación limitada.	Menciona el tema sin profundizar en la motivación o relevancia.	Carece de un propósito o justificación claros.
Metodología		Describe con detalle todas las fases, algoritmos y procedimientos usados.	Menciona correctamente la mayoría de pasos.	Omite partes clave o las explica superficialmente.	La metodología es confusa o incompleta.
Resultados y Análisis	y	Muestra resultados claros y gráficos explicativos. Interpreta y compara adecuadamente los métodos.	Presenta resultados pero con análisis limitado.	Los resultados están incompletos o mal interpretados.	No presenta resultados o son erróneos.
Conclusiones y Trabajo Futuro	y	Resume los hallazgos principales, identifica limitaciones y propone mejoras.	Presenta conclusiones generales.	Conclusiones vagas o sin relación con resultados.	No incluye conclusiones.
Estructura y Redacción		Documento claro, coherente y bien presentado.	Redacción comprensible con leves errores.	Algunos errores formales o de organización.	Redacción deficiente y formato inadecuado.

Cuadro 1: Rúbrica para la evaluación del informe del proyecto Fashion-MNIST.