



5230584518348292231/2998952348759043987308476732179475617/189725005310
0402766 15 24 502948641 73 3375 73 30534/03106
4371141 89226 87 0325010 8540 386964000 1913020 6411 63 521 0/15217352
1271719 33250 72 90785/15 8517 757473073 1020907 4645 57 0382 160344019
977169 60339 44 0135955 414625797 270 78 081 6300 341921240
2388136 17397 14 6157229 64 0791524719699431 374 12 929 1597 094047147
1800640 01635 32 6844413 806 180983642574191 1715 7715 546 6253031166
3692333 60 7249 9253 290 1768 3973 59066198069
174601791270281913071/2970678081118 77634229343273454969451/4323741969
4017211005519079931368017320868140 3710828158660319620838985832239901
35489387030915222098117458681656 5231145235835079292452199952412
567332615666438774376377771881353 7116299856971849532043434627373629
30691892142587974509557101090660952 547414559054821368077940465962703
6700154826779446456512731422194212 828489876942215570852167623454684
1644044275944429948005849502026932790465275857646892878966880527270327

El reconocimiento óptico de caracteres (OCR, por sus siglas en inglés) es el proceso por el cual se traducen o convierten imágenes de dígitos o caracteres (sean éstos manuscritos o de alguna tipografía especial) a un formato representable en nuestra computadora (por ejemplo, ASCII). Esta tarea puede ser más sencilla (por ejemplo, cuando tratamos de determinar el texto escrito en una versión escaneada a buena resolución de un libro) o tornarse casi imposible (recetas indescifrables de médicos, algunos parciales manuscritos de alumnos de métodos numéricos, etc).

El objetivo del trabajo práctico es implementar un método de reconocimiento de dígitos manuscritos basado en la descomposición en valores singulares, y analizar empíricamente los parámetros principales del método.

Como instancias de entrenamiento, se tiene un conjunto de n imágenes de dígitos manuscritos en escala de grises del mismo tamaño y resolución (varias imágenes de cada dígito). Cada una de estas imágenes sabemos a qué dígito se corresponde. En este trabajo consideraremos la popular base de datos MNIST, utilizada como referencia en esta área de investigación¹.

Para $i = 1, \dots, n$, sea $x_i \in \mathbb{R}^m$ la i -ésima imagen de nuestra base de datos almacenada por filas en un vector, y sea $\mu = (x_1 + \dots + x_n)/n$ el promedio de las imágenes. Definimos $X \in \mathbb{R}^{n \times m}$ como la matriz que contiene en la i -ésima fila al vector $(x_i - \mu)^t / \sqrt{n-1}$, y

$$X = U\Sigma V^t$$

a su descomposición en valores singulares, con $U \in \mathbb{R}^{n \times n}$ y $V \in \mathbb{R}^{m \times m}$ matrices ortogonales, y $\Sigma \in \mathbb{R}^{n \times m}$ la matriz diagonal conteniendo en la posición (i, i) al i -ésimo valor singular σ_i . Siendo v_i la columna i de V , definimos para $i = 1, \dots, n$ la *transformación característica* del dígito x_i como el vector $\mathbf{tc}(x_i) = (v_1^t x_i, v_2^t x_i, \dots, v_k^t x_i) \in \mathbb{R}^k$, donde $k \in \{1, \dots, m\}$ es un parámetro de la implementación. Este proceso corresponde a extraer las k primeras *componentes principales* de cada imagen. La intención es que $\mathbf{tc}(x_i)$ resuma la información más relevante de la imagen, descartando los detalles o las zonas que no aportan rasgos distintivos.

Dada una nueva imagen x de un dígito manuscrito, que no se encuentra en el conjunto

¹<http://yann.lecun.com/exdb/mnist/>

inicial de imágenes de entrenamiento, el problema de reconocimiento consiste en determinar a qué dígito corresponde. Para esto, se calcula $\mathbf{tc}(x)$ y se compara con $\mathbf{tc}(x_i)$, para $i = 1, \dots, n$.

Enunciado

Se pide implementar un programa que lea desde archivos las imágenes de entrenamiento de distintos dígitos manuscritos y que, utilizando la descomposición en valores singulares, se calcule la transformación característica de acuerdo con la descripción anterior. Para ello se deberá implementar algún método de estimación de autovalores/autovectores. Dada una nueva imagen de un dígito manuscrito, el programa deberá determinar a qué dígito corresponde. El formato de los archivos de entrada y salida queda a elección del grupo. Si no usan un entorno de desarrollo que incluya bibliotecas para la lectura de archivos de imágenes, sugerimos que utilicen imágenes en formato RAW.

Se deberán realizar experimentos para medir la efectividad del reconocimiento, analizando tanto la influencia de la cantidad k de componentes principales seleccionadas como la influencia de la precisión en el cálculo de los autovalores.

5230584518348292231/2998952348759043987308476732179475617/189725005310

Fecha de entrega:

- *Formato electrónico:* viernes 21 de junio de 2013, hasta las 23:59 hs., enviando el trabajo (informe+código) a metnum.lab@gmail.com. El subject del email debe comenzar con el texto [TP3] seguido de la lista de apellidos de los integrantes del grupo.
- *Formato físico:* lunes 24 de junio de 2013, de 18 a 20hs (en la clase de la práctica).