## CMPE 544 Pattern Recognition

## Assignment 1

by Arda Melikşah Demir

## Support Vector Machine

During the implementation of the support vector machine, linear kernel was used and quadratic problems are solved using CVXOPT.

### Kernel – Weighted Hessian

P = cvxopt.matrix(np.outer(y, y) \* K)

### Linear Term

q = cvxopt.matrix(-np.ones(n\_samples))

### Equality Constraint

A = cvxopt.matrix(y.reshape(1, -1), tc='d')

b = cvxopt.matrix(0.0)

### Box Constraints

if C == 0:

G = cvxopt.matrix(-np.eye(n))

h = cvxopt.matrix(np.zeros(n))

else:

G = cvxopt.matrix(np.vstack([-np.eye(n), np.eye(n)]))

h = cvxopt.matrix(np.hstack([np.zeros(n), np.ones(n)\*C]))

PCA 32

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (Minute) |
| 0.1 | 0.92 | 0.9156 | 0.93 | 0.9213 | 0.9175 | 0.9209 | 2.81 |
| 0.125 | 0.9194 | 0.9163 | 0.93 | 0.9206 | 0.9181 | 0.9209 | 2.94 |
| 0.15 | 0.9194 | 0.9156 | 0.9306 | 0.9213 | 0.9181 | 0.9210 | 3.05 |
| 0.175 | 0.9194 | 0.9163 | 0.93 | 0.9213 | 0.9181 | 0.9210 | 3.15 |
| 0.2 | 0.9181 | 0.9156 | 0.9306 | 0.9206 | 0.9181 | 0.9206 | 3.23 |

PCA = 64

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (Minute) |
| 0.1 | 0.9219 | 0.9231 | 0.9163 | 0.9125 | 0.9231 | 0.9194 | 3.98 |
| 0.125 | 0.9219 | 0.9225 | 0.9175 | 0.9137 | 0.9231 | 0.9197 | 4.16 |
| 0.15 | 0.9219 | 0.9225 | 0.9181 | 0.9125 | 0.9231 | 0.9196 | 4.32 |
| 0.175 | 0.9219 | 0.9231 | 0.9181 | 0.9119 | 0.9231 | 0.9196 | 4.45 |
| 0.2 | 0.9219 | 0.9231 | 0.9181 | 0.9119 | 0.9237 | 0.9197 | 4.57 |

PCA = 128

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (Minute) |
| 0.1 | 0.9113 | 0.9231 | 0.9225 | 0.9169 | 0.9150 | 0.9178 | 5.63 |
| 0.125 | 0.9125 | 0.9231 | 0.9219 | 0.9163 | 0.9150 | 0.9178 | 5.88 |
| 0.15 | 0.9113 | 0.9231 | 0.9219 | 0.9169 | 0.9150 | 0.9176 | 6.10 |
| 0.175 | 0.9125 | 0.9231 | 0.9213 | 0.9181 | 0.9150 | 0.9180 | 6.29 |
| 0.2 | 0.9113 | 0.9225 | 0.9213 | 0.9194 | 0.9150 | 0.9179 | 6.46 |

PCA = 256

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (Minute) |
| 0.1 | 0.9163 | 0.9237 | 0.9137 | 0.9169 | 0.9094 | 0.9160 | 7.96 |
| 0.125 | 0.9163 | 0.9213 | 0.9150 | 0.9169 | 0.9106 | 0.9160 | 8.32 |
| 0.15 | 0.9156 | 0.9225 | 0.9131 | 0.9150 | 0.9081 | 0.9149 | 8.63 |
| 0.175 | 0.9156 | 0.9231 | 0.9119 | 0.9163 | 0.9050 | 0.9144 | 8.90 |
| 0.2 | 0.9156 | 0.9213 | 0.9144 | 0.9156 | 0.9056 | 0.9145 | 9.14 |

## Sklearn SVM Library

### Feature Set Normalized Dataset

### Kernel: Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9187 | 0.905 | 0.9075 | 0.9056 | |  | | --- | | 0.9081 | | 0.909 | 28.3 |
| 0.125 | 0.915 | 0.9056 | 0.91 | 0.905 | 0.9081 | 0.9088 | 28.9 |
| 0.15 | 0.9163 | 0.9038 | 0.9075 | 0.905 | 0.9062 | 0.9078 | 29.8 |
| 0.175 | 0.9169 | 0.9019 | 0.905 | 0.9044 | 0.9069 | 0.908 | 30.2 |
| 0.2 | 0.9194 | 0.9019 | 0.905 | 0.9038 | 0.9069 | 0.9074 | 31.3 |

### Kernel: Poly

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9469 | 0.9425 | 0.9413 | 0.9413 | 0.9431 | 0.943 | 26.0 |
| 0.125 | 0.9481 | 0.9437 | 0.9431 | 0.9400 | 0.9425 | 0.9435 | 26.7 |
| 0.15 | 0.9494 | 0.9456 | 0.9419 | 0.9406 | 0.9450 | 0.9445 | 26.3 |
| 0.175 | 0.9494 | 0.9463 | 0.9425 | 0.9394 | 0.945 | 0.9445 | 27.0 |
| 0.2 | 0.95 | 0.9431 | 0.9419 | 0.9381 | 0.945 | 0.9436 | 27.1 |

### Kernel: RBF

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.935 | 0.9344 | 0.9325 | 0.9337 | 0.9387 | 0.9349 | 51.9 |
| 0.125 | 0.9369 | 0.9360 | 0.9331 | 0.9344 | 0.9387 | 0.936 | 48.4 |
| 0.15 | 0.9381 | 0.9375 | 0.9337 | 0.9331 | 0.94 | 0.9365 | 45.6 |
| 0.175 | 0.94 | 0.9375 | 0.9363 | 0.935 | 0.9419 | 0.9381 | 44.4 |
| 0.2 | 0.9413 | 0.9381 | 0.9381 | 0.935 | 0.9413 | 0.9387 | 42.7 |

### Feature Set Binary İmages

### Kernel: Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.91 | 0.8988 | 0.8994 | 0.9044 | 0.9156 | 0.9056 | 30.4 |
| 0.125 | 0.9062 | 0.8969 | 0.8981 | 0.905 | 0.9131 | 0.9039 | 31.8 |
| 0.15 | 0.9075 | 0.8969 | 0.8975 | 0.9075 | 0.9137 | 0.9046 | 33.2 |
| 0.175 | 0.9069 | 0.8962 | 0.8994 | 0.9062 | 0.9144 | 0.9046 | 34.0 |
| 0.2 | 0.9069 | 0.8969 | 0.8975 | 0.9044 | 0.9113 | 0.9034 | 35.0 |

### Kernel: Poly

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9419 | 0.935 | 0.9344 | 0.9419 | 0.9456 | 0.9398 | 26.5 |
| 0.125 | 0.9413 | 0.9375 | 0.9331 | 0.9437 | 0.945 | 0.9401 | 27.4 |
| 0.15 | 0.9437 | 0.9344 | 0.9337 | 0.9431 | 0.945 | 0.94 | 27.5 |
| 0.175 | 0.9437 | 0.9363 | 0.9331 | 0.9437 | 0.9456 | 0.9405 | 30.9 |
| 0.2 | 0.9437 | 0.9369 | 0.9325 | 0.9444 | 0.9444 | 0.9404 | 28.9 |

### Kernel: RBF

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9369 | 0.9194 | 0.93 | 0.9294 | 0.9363 | 0.9304 | 55.1 |
| 0.125 | 0.9381 | 0.9213 | 0.9294 | 0.9306 | 0.9381 | 0.9315 | 51.6 |
| 0.15 | 0.9387 | 0.9225 | 0.9294 | 0.9306 | 0.9381 | 0.9319 | 49.8 |
| 0.175 | 0.94 | 0.9244 | 0.9281 | 0.9306 | 0.9369 | 0.932 | 47.1 |
| 0.2 | 0.9413 | 0.9263 | 0.9287 | 0.9331 | 0.9375 | 0.9334 | 46.7 |

### Feature Set PCA-32

### Kernel: Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9294 | 0.9081 | 0.9237 | 0.9263 | 0.92 | 0.9215 | 2.6 |
| 0.125 | 0.9294 | 0.9087 | 0.9237 | 0.9263 | 0.9213 | 0.9219 | 2.6 |
| 0.15 | 0.9294 | 0.9094 | 0.9237 | 0.9256 | 0.9213 | 0.9219 | 2.8 |
| 0.175 | 0.9294 | 0.9094 | 0.9237 | 0.9256 | 0.9213 | 0.9219 | 2.9 |
| 0.2 | 0.9294 | 0.9094 | 0.9237 | 0.9256 | 0.9213 | 0.9219 | 3.1 |

### Kernel: Poly

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9531 | 0.9337 | 0.9931 | 0.9494 | 0.9469 | 0.9432 | 2.1 |
| 0.125 | 0.9537 | 0.9344 | 0.9337 | 0.9494 | 0.9469 | 0.9436 | 2.1 |
| 0.15 | 0.9531 | 0.9344 | 0.9356 | 0.9494 | 0.9475 | 0.944 | 2.1 |
| 0.175 | 0.9537 | 0.9356 | 0.9363 | 0.9487 | 0.9475 | 0.9444 | 2.0 |
| 0.2 | 0.9531 | 0.9344 | 0.9375 | 0.9506 | 0.9487 | 0.9449 | 2.0 |

### Kernel: RBF

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9519 | 0.9319 | 0.9313 | 0.9413 | 0.9444 | 0.9401 | 5.3 |
| 0.125 | 0.9519 | 0.9344 | 0.9325 | 0.9425 | 0.9450 | 0.9412 | 4.9 |
| 0.15 | 0.9506 | 0.9337 | 0.9337 | 0.9431 | 0.9463 | 0.9415 | 4.7 |
| 0.175 | 0.9506 | 0.9344 | 0.9337 | 0.945 | 0.9444 | 0.9416 | 4.5 |
| 0.2 | 0.9519 | 0.9344 | 0.9344 | 0.945 | 0.9444 | 0.9420 | 4.4 |

### Feature Set PCA-64

### Kernel: Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9244 | 0.9281 | 0.9244 | 0.9144 | 0.92 | 0.9223 | 3.3 |
| 0.125 | 0.9237 | 0.9281 | 0.9256 | 0.9137 | 0.9206 | 0.9224 | 3.6 |
| 0.15 | 0.9237 | 0.9281 | 0.9256 | 0.9137 | 0.9213 | 0.9225 | 3.8 |
| 0.175 | 0.9231 | 0.9287 | 0.9263 | 0.9131 | 0.9213 | 0.9225 | 4.0 |
| 0.2 | 0.9231 | 0.9287 | 0.9263 | 0.915 | 0.9213 | 0.9229 | 4.1 |

### Kernel: Poly

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9469 | 0.9463 | 0.9481 | 0.9325 | 0.9406 | 0.9429 | 2.8 |
| 0.125 | 0.9487 | 0.9475 | 0.9487 | 0.9344 | 0.9413 | 0.9441 | 2.8 |
| 0.15 | 0.9481 | 0.9469 | 0.9487 | 0.9375 | 0.9425 | 0.9447 | 2.8 |
| 0.175 | 0.9487 | 0.9487 | 0.9487 | 0.9394 | 0.9419 | 0.9455 | 2.7 |
| 0.2 | 0.9481 | 0.9494 | 0.9494 | 0.9394 | 0.9413 | 0.9455 | 2.7 |

### Kernel: RBF

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9387 | 0.9469 | 0.9400 | 0.9306 | 0.9381 | 0.9389 | 7.2 |
| 0.125 | 0.9413 | 0.9475 | 0.9413 | 0.9319 | 0.9381 | 0.94 | 6.4 |
| 0.15 | 0.9406 | 0.9581 | 0.9419 | 0.9331 | 0.9387 | 0.9405 | 6.1 |
| 0.175 | 0.9406 | 0.9487 | 0.9425 | 0.9337 | 0.9400 | 0.9411 | 5.8 |
| 0.2 | 0.9406 | 0.9494 | 0.9431 | 0.9425 | 0.9400 | 0.9411 | 5.7 |

### Feature Set PCA-128

### Kernel: Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9231 | 0.9263 | 0.9206 | 0.9206 | 0.9200 | 0.9221 | 6.0 |
| 0.125 | 0.9244 | 0.9263 | 0.9194 | 0.9212 | 0.9206 | 0.9224 | 6.3 |
| 0.15 | 0.9231 | 0.9256 | 0.9200 | 0.9206 | 0.9200 | 0.9219 | 6.8 |
| 0.175 | 0.9213 | 0.9263 | 0.9200 | 0.9200 | 0.9194 | 0.9214 | 7.4 |
| 0.2 | 0.9213 | 0.9244 | 0.9200 | 0.9194 | 0.9194 | 0.9209 | 7.8 |

### Kernel: Poly

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9437 | 0.9425 | 0.9425 | 0.9419 | 0.9450 | 0.9431 | 5.1 |
| 0.125 | 0.9481 | 0.9431 | 0.9419 | 0.9437 | 0.9469 | 0.9447 | 5.0 |
| 0.15 | 0.9487 | 0.9437 | 0.9444 | 0.9437 | 0.9469 | 0.9455 | 4.8 |
| 0.175 | 0.9494 | 0.9444 | 0.9437 | 0.9450 | 0.9463 | 0.9457 | 4.8 |
| 0.2 | 0.9494 | 0.9463 | 0.9431 | 0.9444 | 0.9469 | 0.9460 | 4.9 |

### Kernel: RBF

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9319 | 0.9406 | 0.9375 | 0.9381 | 0.9356 | 0.9367 | 11.5 |
| 0.125 | 0.9331 | 0.9413 | 0.9394 | 0.9387 | 0.9363 | 0.9378 | 10.9 |
| 0.15 | 0.9350 | 0.9437 | 0.9387 | 0.9394 | 0.9356 | 0.9385 | 10.6 |
| 0.175 | 0.9381 | 0.9437 | 0.9394 | 0.9400 | 0.9369 | 0.9396 | 9.6 |
| 0.2 | 0.9394 | 0.9437 | 0.9413 | 0.9413 | 0.9381 | 0.9408 | 9.4 |

### Feature Set PCA-256

### Kernel: Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9131 | 0.9206 | 0.9231 | 0.9150 | 0.9256 | 0.9195 | 11.6 |
| 0.125 | 0.9119 | 0.9219 | 0.9213 | 0.9131 | 0.9250 | 0.9186 | 12.1 |
| 0.15 | 0.9106 | 0.9206 | 0.9213 | 0.9144 | 0.9244 | 0.9183 | 12.7 |
| 0.175 | 0.9313 | 0.9206 | 0.9206 | 0.9144 | 0.9237 | 0.9181 | 13.4 |
| 0.2 | 0.9119 | 0.9206 | 0.9213 | 0.9144 | 0.9231 | 0.9183 | 13.6 |

### Kernel: Poly

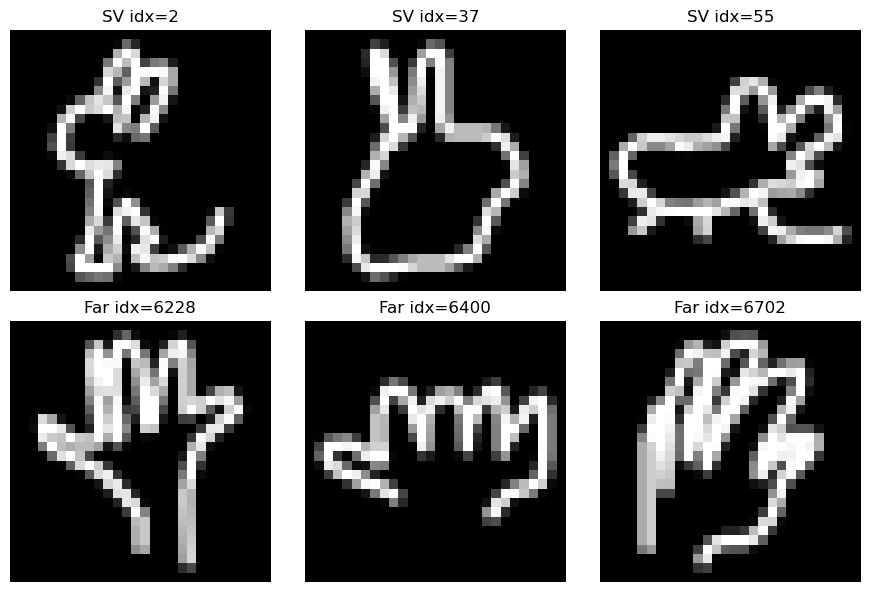
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9313 | 0.9419 | 0.9463 | 0.9400 | 0.9475 | 0.9414 | 10.0 |
| 0.125 | 0.9325 | 0.9444 | 0.9463 | 0.9400 | 0.9475 | 0.9421 | 9.8 |
| 0.15 | 0.9337 | 0.9450 | 0.9470 | 0.9250 | 0.9475 | 0.9432 | 9.7 |
| 0.175 | 0.9344 | 0.9463 | 0.9475 | 0.9431 | 0.9456 | 0.9434 | 9.3 |
| 0.2 | 0.9337 | 0.9481 | 0.9444 | 0.9431 | 0.9469 | 0.9433 | 9.3 |

### Kernel: RBF

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Avg Accuracy | Time (second) |
| 0.1 | 0.9281 | 0.9356 | 0.9363 | 0.9300 | 0.9437 | 0.9347 | 33.3 |
| 0.125 | 0.9300 | 0.9356 | 0.9400 | 0.9300 | 0.9431 | 0.9357 | 31.7 |
| 0.15 | 0.9319 | 0.9363 | 0.9406 | 0.9300 | 0.9450 | 0.9367 | 30.1 |
| 0.175 | 0.9331 | 0.9375 | 0.9419 | 0.9319 | 0.9463 | 0.9381 | 28.9 |
| 0.2 | 0.9331 | 0.9387 | 0.9425 | 0.9331 | 0.9456 | 0.9386 | 27.5 |

In the end, linear kernels for SVM implementations yielded similar results. It was also observed that poly and RBF kernels returned higher accuracies than linear with the polynomial kernel being the highest one. Other than that, differently preprocessed train set didn’t change the accuracy in a significant matter. However, PCA 128-dimensional data still returned the best results.

### Support Vectors



When we examine the support vectors, we can see that support vectors are very ambiguous and don’t look like anything. On the other hand, far from margin vectors clearly represent the label.

## Clustering

### Implementation

In this project, K-Means algorithm is implemented with k = 5 and the dataset is clustered. Initially during the implementation of K-Means algorithm, K-Means++ wasn’t used and all the original centroids were selected randomly. This resulted in not favorable results where all of accuracy metrics were between 0.4 and 0.6.

Later KMeansClustering++ algorithm was used to initialize the centroids. While it definitely improved the accuracy of estimations, it still didn’t yield very good results.

2 different datasets with 3 different distance metrics is used while testing the clustering algorithm. Those distance metrics are Manhattan, Euclidean and Cosine distances. While the datasets used in the task are normalized images and PCA projected 128 dimensional image vectors. The number of 128 is chosen due to high yielding accuracy results in the same dataset in different algorithms.

While choosing 2 additional metrics to evaluate the Clustering algorithm, Calinski Harabasz and Davies-Bouldin indexes were chosen.

Davies-Bouldin Index shows how good the clustering in the dataset is. It measures the compactness of each cluster and how far the clusters are. Lower Davies-Bouldin Index means better and clearer clusters while the higher means vice-versa. It is calculated by:

Another metric that is used is Calinski-Harabasz Index. It is very similar to DB index in the sense that it measures how good the clusters are in a dataset. It measures how close the points and how far the clusters.

B is the sum of squares between clusters.

W is the sum of squares within clusters.

N is the total number of data points.

K is the number of clusters.

### Euclidean Distance Clustering With Accuracy and SSE

|  |  |  |
| --- | --- | --- |
| 784-Dimensional Normalized Data | Regular KMeans | KMeans++ |
| Accuracy | 0.51 | 0.57 |
| SSE | 1334888 | 1274285 |
| CHI | 381 | 637 |
| DBI | 5.7 | 4.47 |

|  |  |  |
| --- | --- | --- |
| 256-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.54 | 0.57 |
| SSE | 1161614 | 1141254 |
| CHI | 597 | 696 |
| DBI | 4.46 | 4.26 |

|  |  |  |
| --- | --- | --- |
| 128-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.53 | 0.57 |
| SSE | 953239 | 934950 |
| CHI | 732 | 844 |
| DBI | 4.43 | 3.86 |

|  |  |  |
| --- | --- | --- |
| 64-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.41 | 0.57 |
| SSE | 719378 | 687985 |
| CHI | 871 | 1139 |
| DBI | 4.08 | 3.32 |

|  |  |  |
| --- | --- | --- |
| 32-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.49 | 0.58 |
| SSE | 482673 | 462222 |
| CHI | 1204 | 1688 |
| DBI | 3.22 | 2.69 |

### Manhattan and Cosine Distance Results

### Manhattan

|  |  |  |
| --- | --- | --- |
| 32-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.54 | 0.53 |
| SSE | 491356 | 468626 |
| CHI | 1291 | 1596 |
| DBI | 3.73 | 2.94 |

|  |  |  |
| --- | --- | --- |
| 64-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.55 | 0.56 |
| SSE | 710044 | 694970 |
| CHI | 949 | 1078 |
| DBI | 3.82 | 3.74 |

|  |  |  |
| --- | --- | --- |
| 128-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.54 | 0.55 |
| SSE | 961022 | 941734 |
| CHI | 685 | 802 |
| DBI | 5.43 | 4.30 |

|  |  |  |
| --- | --- | --- |
| 256-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.51 | 0.56 |
| SSE | 1183111 | 1147628 |
| CHI | 495 | 665 |
| DBI | 5.00 | 4.74 |

|  |  |  |
| --- | --- | --- |
| Normalized Data Full Dimensional | Regular KMeans | KMeans++ |
| Accuracy | 0.48 | 0.51 |
| SSE | 1295943 | 1301650 |
| CHI | 542 | 518 |
| DBI | 5.95 | 3.76 |

### Cosine

Cosine Distance computes the given formula to compute the distance between two data points:

|  |  |  |
| --- | --- | --- |
| 32-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.46 | 0.57 |
| SSE | 492838 | 462769 |
| CHI | 1291 | 1596 |
| DBI | 3.12 | 2.73 |

|  |  |  |
| --- | --- | --- |
| 64-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.53 | 0.57 |
| SSE | 705032 | 688477 |
| CHI | 949 | 1078 |
| DBI | 4.06 | 3.35 |

|  |  |  |
| --- | --- | --- |
| 128-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.39 | 0.57 |
| SSE | 962175 | 935455 |
| CHI | 685 | 802 |
| DBI | 4.82 | 3.89 |

|  |  |  |
| --- | --- | --- |
| 256-Dimensional Data (Normalized) | Regular KMeans | KMeans++ |
| Accuracy | 0.51 | 0.57 |
| SSE | 1184262 | 1141740 |
| CHI | 495 | 665 |
| DBI | 5.03 | 4.29 |

|  |  |  |
| --- | --- | --- |
| Normalized Data Full Dimensional | Regular KMeans | KMeans++ |
| Accuracy | 0.47 | 0.57 |
| SSE | 1310605 | 1274753 |
| CHI | 481 | 634 |
| DBI | 5.15 | 4.51 |