PCA-Based KNN Animal Classification

CSE6363 – Machine Learning
Final Project

Bardia Mojra
May 14, 2021
CSE6363 – Dr. Huber
University of Texas at Arlington

Introduction

- Principal Component Analysis (PCA)
  - Feature extraction
  - Data compression
- K-Nearest Neighbor (KNN)
  - Classification

PCA

- Process dataset:
  - Training and test sets must be consistent
  - Resize and crop square
  - Reshape images to row vectors and stack dataset
- Normalize dataset, standardize
- Calculate Covariance, EigenVectors and EigenValues:
  - Sort EigenVectors by EigenValues
- Dot product multiply top n EigenVectors (PC’s) by the dataset

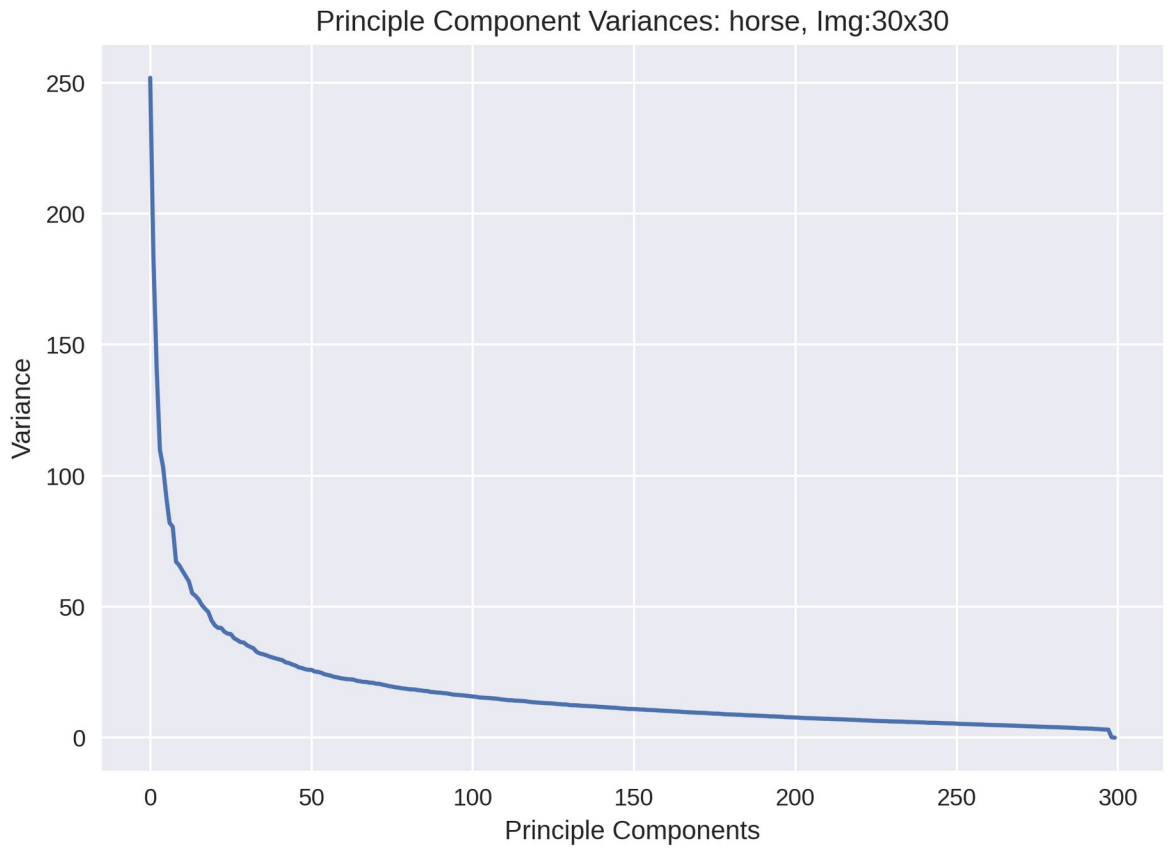
KNN

- KNN classification
- Simple voting and weighted voting (distance)

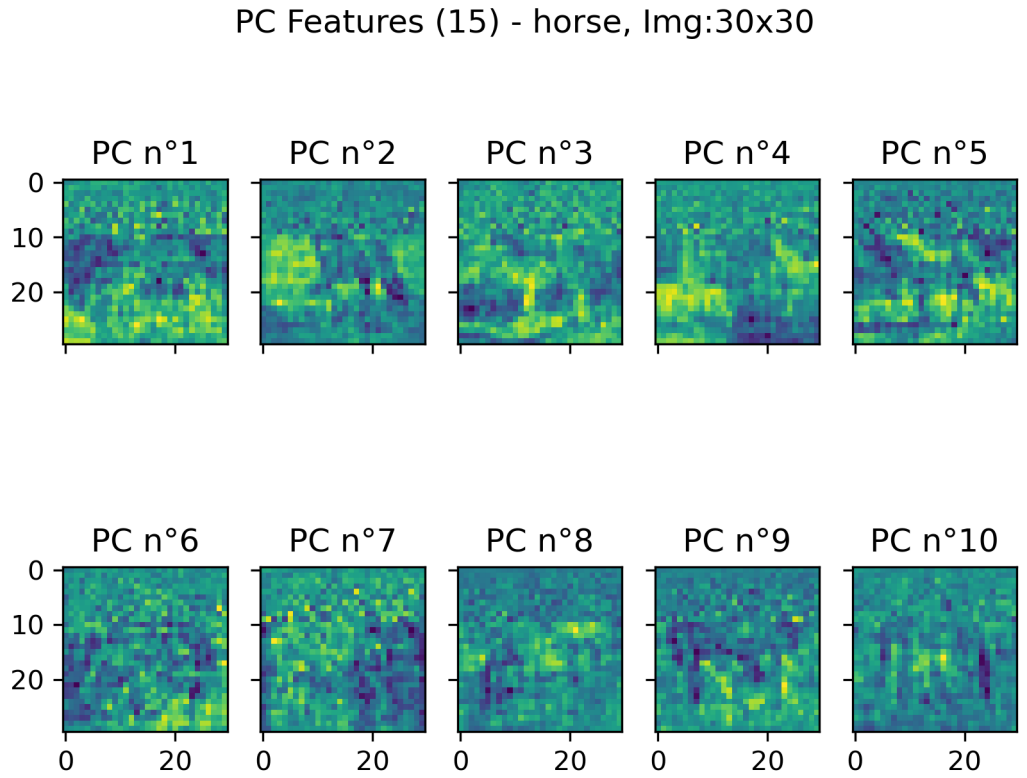
Test Configuration

- knn_k = 7
- image_size = 30 # pixels, equal height and width
- samples_per_class = 300
- shuffles = 10 # shuffle n times to mix data
- testfrac = .2 # 0-1.0 | test set fraction of main set
- PC_ratio = 100 # % / 5, 10, 30, 60, 100%
- numPC = int((PC_ratio/100)*(image_size**2))
- Clen = 4 # dog, horse, chick, sheep

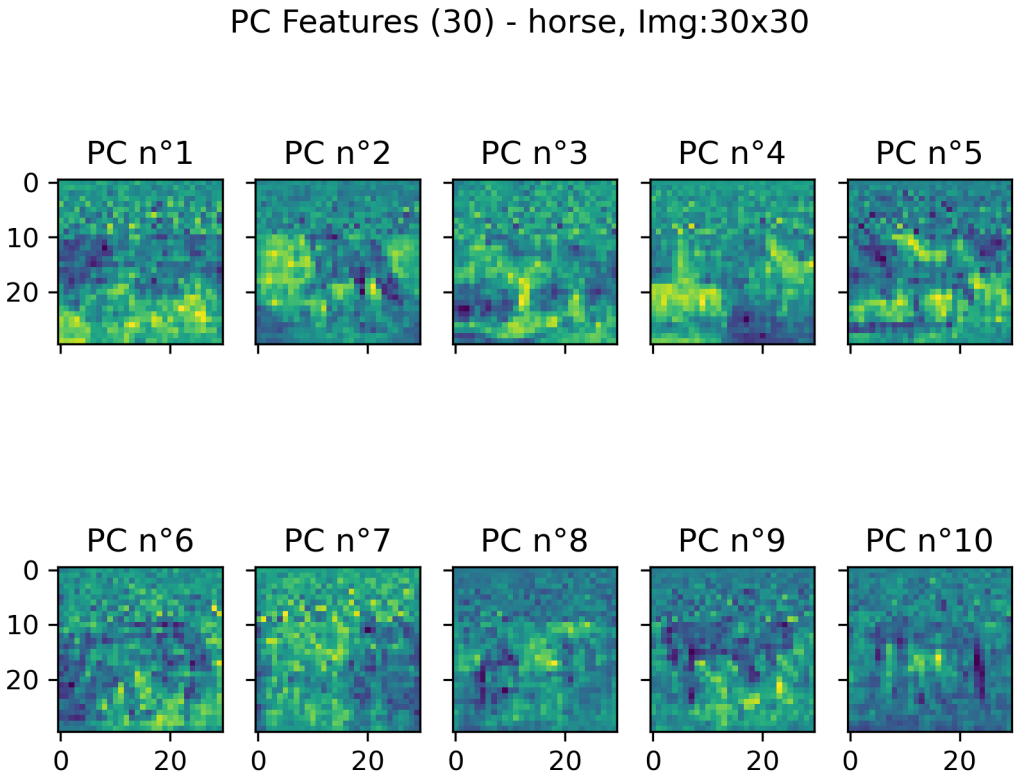
Principle Components and Covariance



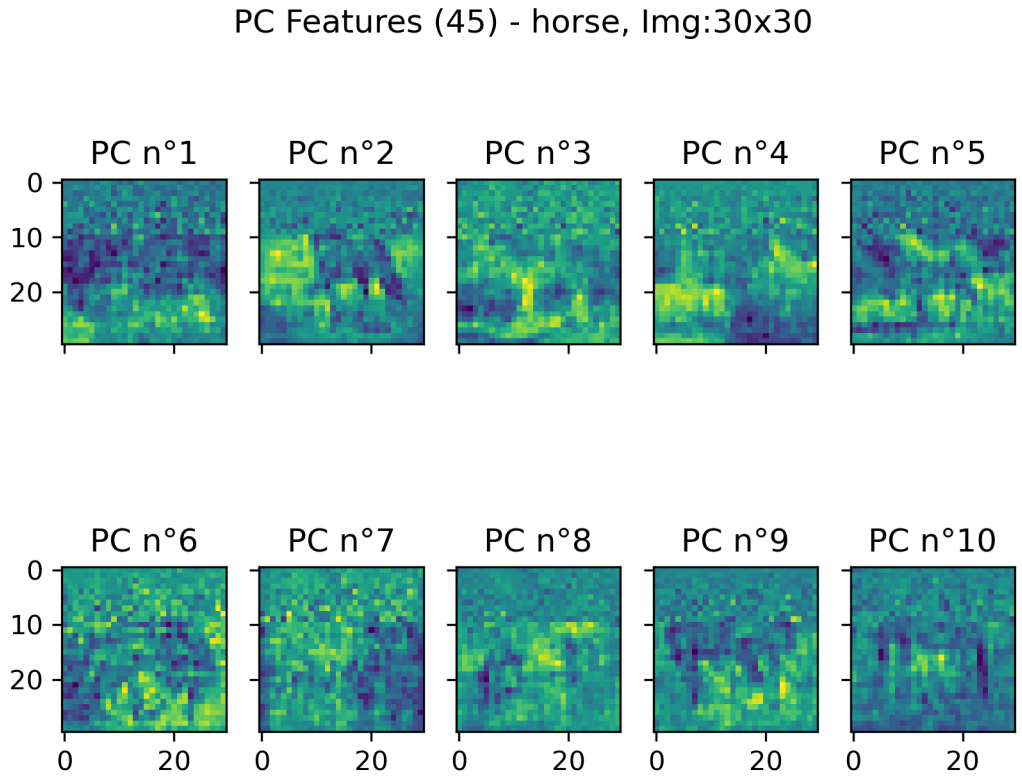
Principal Components



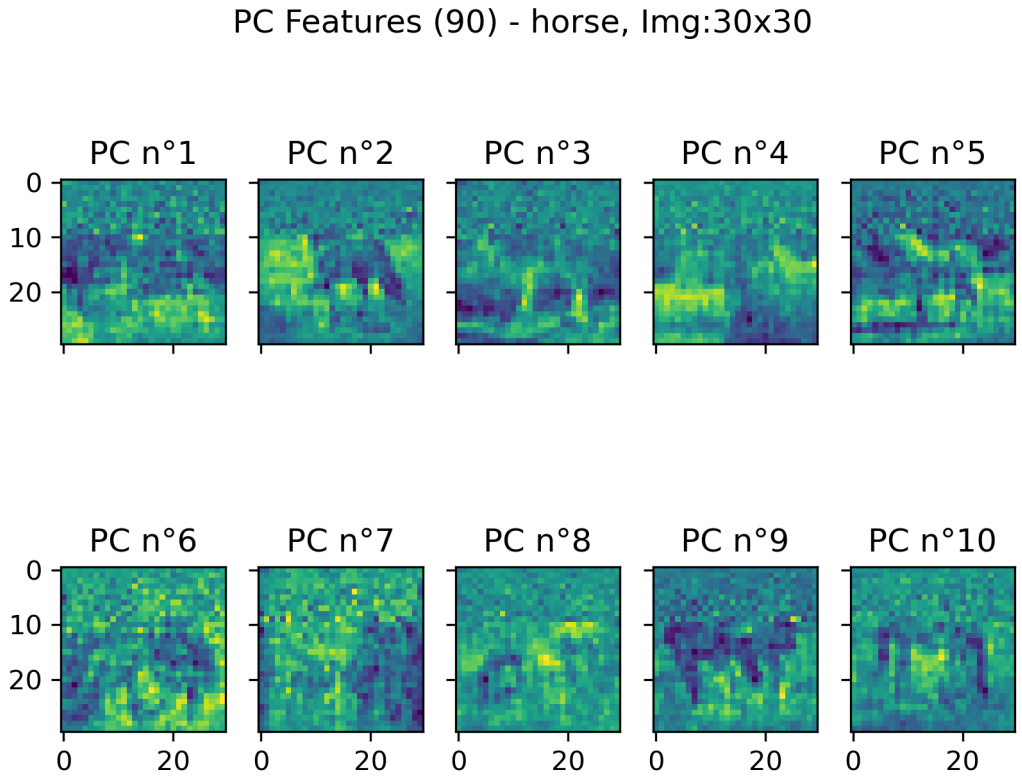
Principal Components



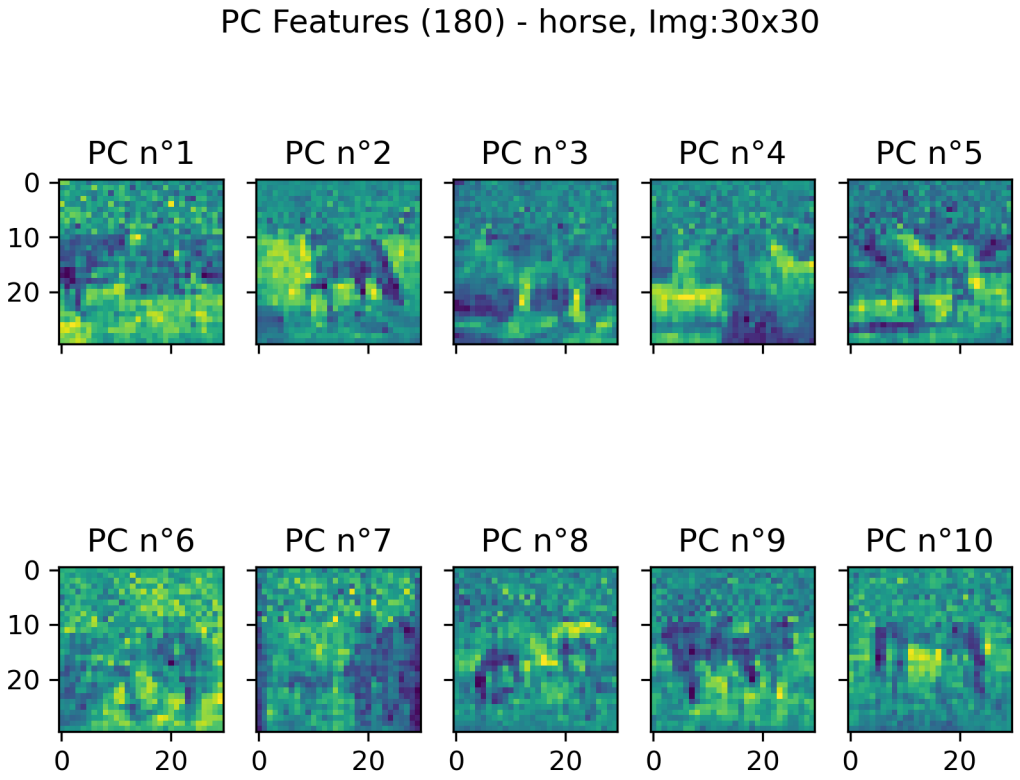
Principal Components



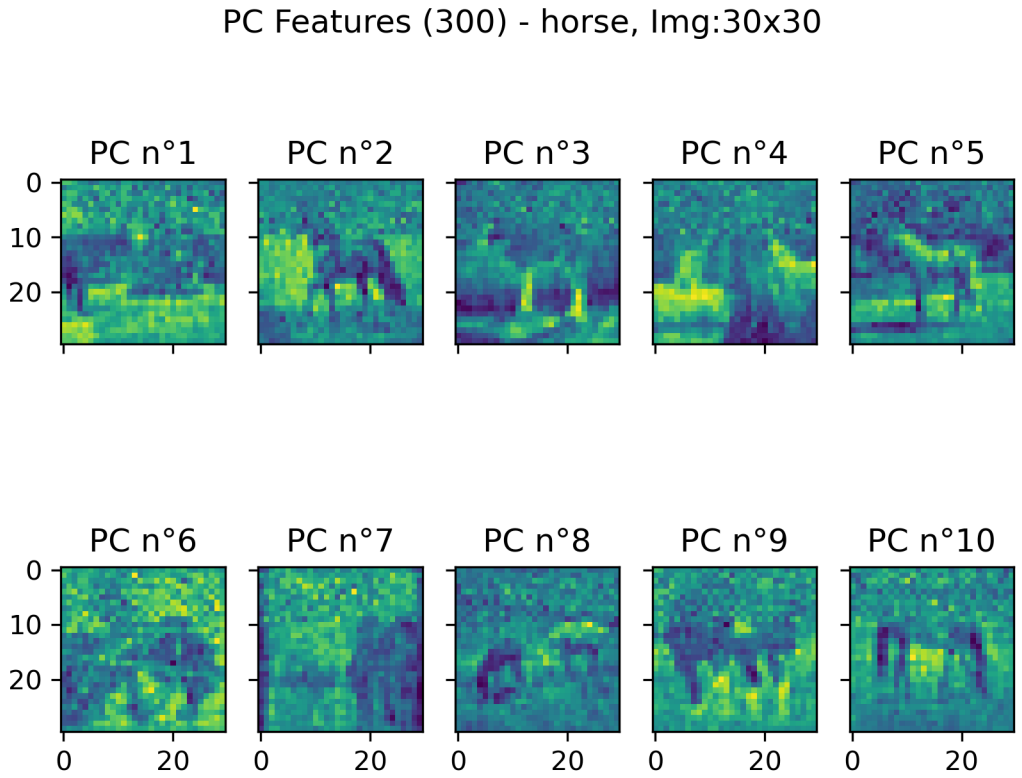
Principal Components



Principal Components



Principal Components



Results

Experiment: CL4_Res48_PC15training data (weighted): 100.000%training data (simple): 87.812%test data (weighted): 85.417%test data (simple): 82.500%

Experiment: CL4_Res30_PC15training data (weighted): 100.000%training data (simple): 89.896%test data (weighted): 82.083%test data (simple): 81.667%

Experiment: CL4_Res48_PC30training data (weighted): 100.000%training data (simple): 91.771%test data (weighted): 87.083%test data (simple): 87.500%

Experiment: CL4_Res30_PC30training data (weighted): 100.000%training data (simple): 93.646%test data (weighted): 91.667%test data (simple): 90.833%

Experiment: CL4_Res30_PC45training data (weighted): 100.000%training data (simple): 94.896%test data (weighted): 93.333%test data (simple): 92.500%

Experiment: CL4_Res48_PC45training data (weighted): 100.000%training data (simple): 93.438%test data (weighted): 93.333%test data (simple): 92.500%

Experiment: CL4_Res48_PC90training data (weighted): 100.000%training data (simple): 94.062%test data (weighted): 89.583%test data (simple): 88.750%

Experiment: CL4_Res30_PC90training data (weighted): 100.000%training data (simple): 94.792%test data (weighted): 91.667%test data (simple): 89.167%

Experiment: CL4_Res30_PC180training data (weighted): 100.000%training data (simple): 95.625%test data (weighted): 93.750%test data (simple): 93.333%

Experiment: CL4_Res30_PC300training data (weighted): 100.000%training data (simple): 96.562%test data (weighted): 90.833%test data (simple): 91.250%

Experiment: CL4_Res48_PC300training data (weighted): 100.000%training data (simple): 91.875%test data (weighted): 85.833%test data (simple): 87.083%

Experiment: CL4_Res48_PC180training data (weighted): 100.000%training data (simple): 89.583% test data (weighted): 85.417%test data (simple): 84.167%

› Thank you!