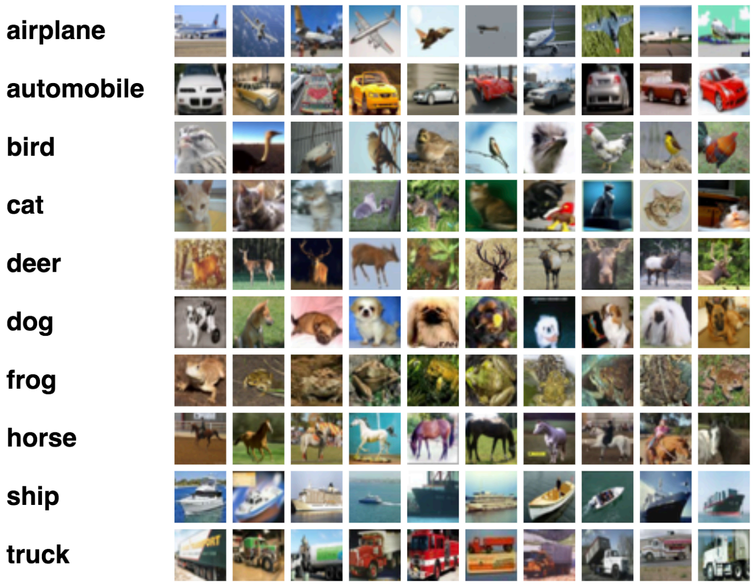
Final Presentation
Machine Learning

Juleana Villegas and Xinda Wu
University of Bonn

July 17, 2018

Data


airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

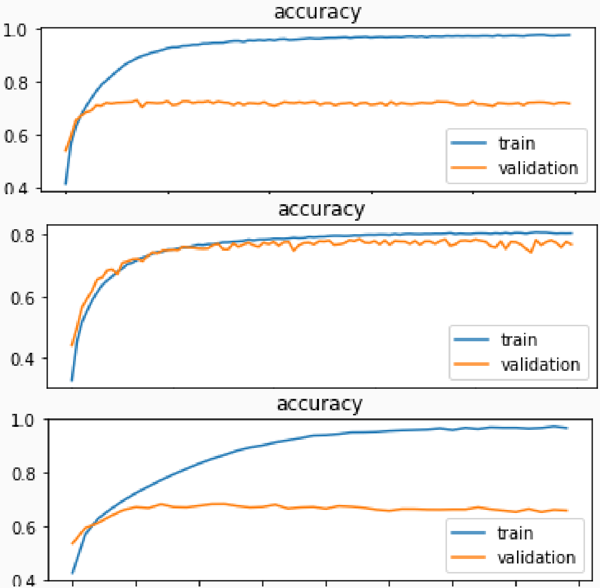
Models

CV(32) → MP(2) → CV(64) → CV(64) → MP(2) → D(512) → D(10)

CV(32) → CV(32) → MP(2) → CV(64) → CV(64) → MP(2) → D(512) → D(10)

CV(16) → MP(2) → CV(32) → MP(2) → CV(64) → MP(2) → D(512) → D(10)

Accuracy


accuracy
1.0
0.8
0.6
0.4
train
validation
accuracy
0.8
0.6
0.4
train
validation
accuracy
1.0
0.8
0.6
0.4
train
validation

Some results

airplane, Pred: airplane

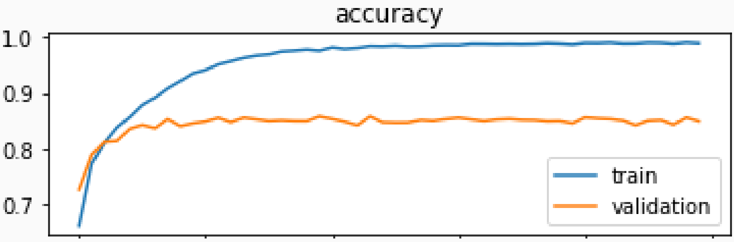
bird, Pred: airplane

bird, Pred: cat

Confusion Matrix

$$
\begin{pmatrix}
747 & 33 & 37 & 15 & 27 & 6 & 7 & 11 & 76 & 41 \\
18 & 868 & 3 & 4 & 3 & 1 & 7 & 2 & 18 & 76 \\
63 & 12 & 596 & 75 & 86 & 54 & 54 & 27 & 23 & 10 \\
28 & 16 & 57 & 503 & 73 & 137 & 70 & 50 & 28 & 38 \\
22 & 5 & 61 & 48 & 690 & 19 & 28 & 98 & 17 & 12 \\
21 & 11 & 53 & 154 & 67 & 595 & 18 & 58 & 9 & 14 \\
8 & 9 & 44 & 66 & 50 & 19 & 766 & 12 & 12 & 14 \\
22 & 10 & 33 & 34 & 53 & 41 & 7 & 772 & 6 & 22 \\
65 & 33 & 10 & 15 & 1 & 3 & 5 & 8 & 817 & 43 \\
31 & 88 & 8 & 10 & 5 & 2 & 4 & 3 & 29 & 820
\end{pmatrix}
$$

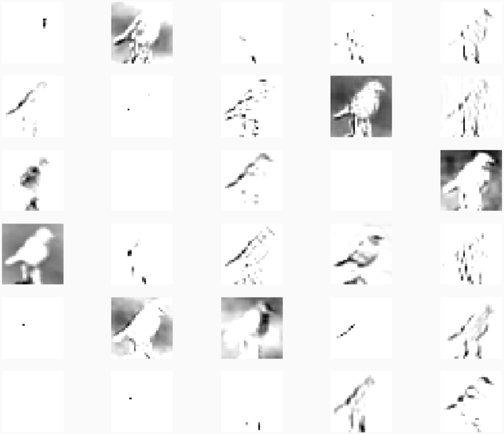
Merge Classes



Intermediate Layer

Intermediate Layer

Intermediate Layer



Principal Component Analysis

Principal component analysis (PCA) is a dimensionality reduction method that convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

Principal Component Analysis

Motivation

1. Computational efficiency
2. Visualization of dataset in high dimension

Principal Component Analysis

Basic ideas

Find an orthogonal transformation mapping the data in high dimension to lower dimension preserving the relevant information.

Principal Component Analysis

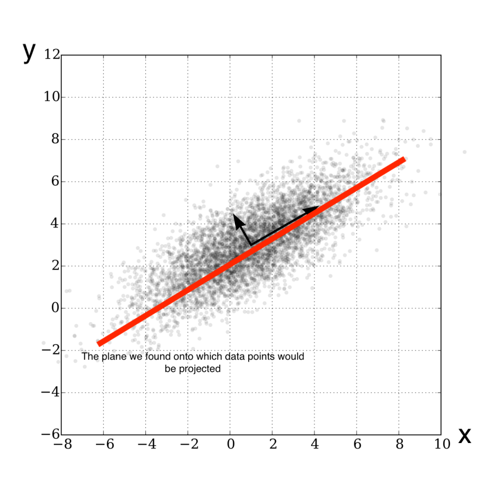


Figure: A 2-dim example(from wikipedia)

Principal Component Analysis

Theorem

Let $\{x_1, \ldots, x_n\}$ *be a set of vectors from* $\mathbb{R}^d$*, and let* $q \in \mathbb{N}$ *with* $q \leq d$*. The minimization problem*

$$\min_{\mu \in \mathbb{R}^d, \lambda_i \in \mathbb{R}^q, V_q \in \mathbb{R}^{d \times q}} \sum_{i=1}^{n} \|x_i - (\mu + V_q \lambda_i)\|^2$$

subject to $V_q^T V_q = I_q$

is solved for $\mu = \bar{x}, \lambda_i = V_q^T(x_i - \bar{x})$*, and* V_q *contains column-wise q orthogonalized unit eigenvectors of* $X^T X$ *for the q largest eigenvalues. (The solution is unique up to rotation)*

Principal Component Analysis

We see that either solve these two basic problems we could get a PCA:

1. Compute the eigenvector of $X^T X$.
2. Compute a singular value of X.

Remark: PCA is only valid for linear case, while in the nonlinear case, that is, the data lies in a general manifold, we have to apply for other methods like diffusion maps.

Another Perspective

The way we derive PCA is based on the problem of least squares error

$$\sum_{i=1}^{n} \|x_i - (\mu + V_q \lambda_i)\|^2$$

with respect to μ, V_q and λ_i. But there is another ***statistics approach*** to interpret this method.

Another Perspective

Actually, for a vector-valued random variable $\hat{X}$ the eigenvectors we got from X^TX based on the previous theorem are those ones of directions on which the data has largest variance. In the order of components,

$$Var[v_1^T\hat{X}] \geq Var[v_2^T\hat{X}] \geq \dots$$

And the eigenvalues Λ_i are equal to the variances of random variables $Y_i := v_i^T\hat{X}$

Decide the number q

Setting a threshold $h \in [0, 1]$, pick q as the smallest number such that

$$\frac{\sum_{i=1}^{q} Var[v_i^T \hat{X}]}{\sum_{i=1}^{d} Var[v_i^T \hat{X}]} = \frac{\sum_{i=1}^{q} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq h$$

Thank you!!