EIGEN VALUES AND EIGEN VECTORS

Your input: find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$

Start from forming a new matrix by subtracting λ from the diagonal entries of the given matrix:

$$\begin{bmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{bmatrix}$$

Find the determinant of the obtained matrix:

$\begin{vmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 3$ (for steps, see determinant calculator)

This is a characteristic polynomial.

Solve the equation $\lambda^2 - 4\lambda + 3 = 0$.

The roots are:

$\lambda_1 = 1$

$\lambda_2 = 3$

These are the eigenvalues.

Next, find the eigenvectors.

a. $\lambda = 1$

$$\begin{bmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix}$$

Perform row operations to obtain the rref of the matrix:

$$\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$ (for steps, see rref calculator)

Now, solve the matrix equation $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

If we take $v_1 = t$, then $v_1 = t$, $v_2 = 0$.

Therefore, $\mathbf{v} = \begin{bmatrix} t \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} t$

b. $\lambda = 3$

$$\begin{bmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix}$$

Perform row operations to obtain the rref of the matrix:

$$\begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$ (for steps, see rref calculator)

Now, solve the matrix equation $\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

If we take $v_2 = t$, then $v_1 = t$, $v_2 = t$.

Therefore, $\mathbf{v} = \begin{bmatrix} t \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} t$

ANSWER

Eigenvalue: 1, **eigenvector:** $\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Eigenvalue: 3, **eigenvector:** $\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Dimensionality Reduction

- Significant improvements can be achieved by first mapping the data into a *lower-dimensional* space.

$$x = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_N \end{bmatrix} \text{---> reduce dimensionality --->} \; y = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_K \end{bmatrix} \quad (K << N)$$

- Dimensionality can be reduced by:
  - Combining features using a linear or non-linear transformations.
  - Selecting a subset of features (i.e., feature selection).

Data Dimensionality

- From a theoretical point of view, increasing the number of features should lead to better performance.

- In practice, the inclusion of more features leads to worse performance (i.e., curse of dimensionality).

- The number of training examples required increases exponentially with dimensionality.

Application of Dimensionality Reduction

- Customer relationship management
- Text mining
- Image retrieval
- Microarray data analysis
- Protein classification
- Face recognition
- Handwritten digit recognition
- Intrusion detection

Feature Selection

- Definition
  - A process that chooses an optimal subset of features according to a objective function
- Objectives
  - To reduce dimensionality and remove noise
  - To improve mining performance
    - Speed of learning
    - Predictive accuracy
    - Simplicity and comprehensibility of mined results

Feature Extraction

- Feature reduction refers to the mapping of the original high-dimensional data onto a lower-dimensional space
- Given a set of data points of p variables $\{x_1, x_2, \cdots, x_n\}$
  Compute their low-dimensional representation
- Criterion for feature reduction can be different based on different problem settings.
  - Unsupervised setting: minimize the information loss
  - Supervised setting: maximize the class discrimination

Feature Reduction vs. Feature Selection

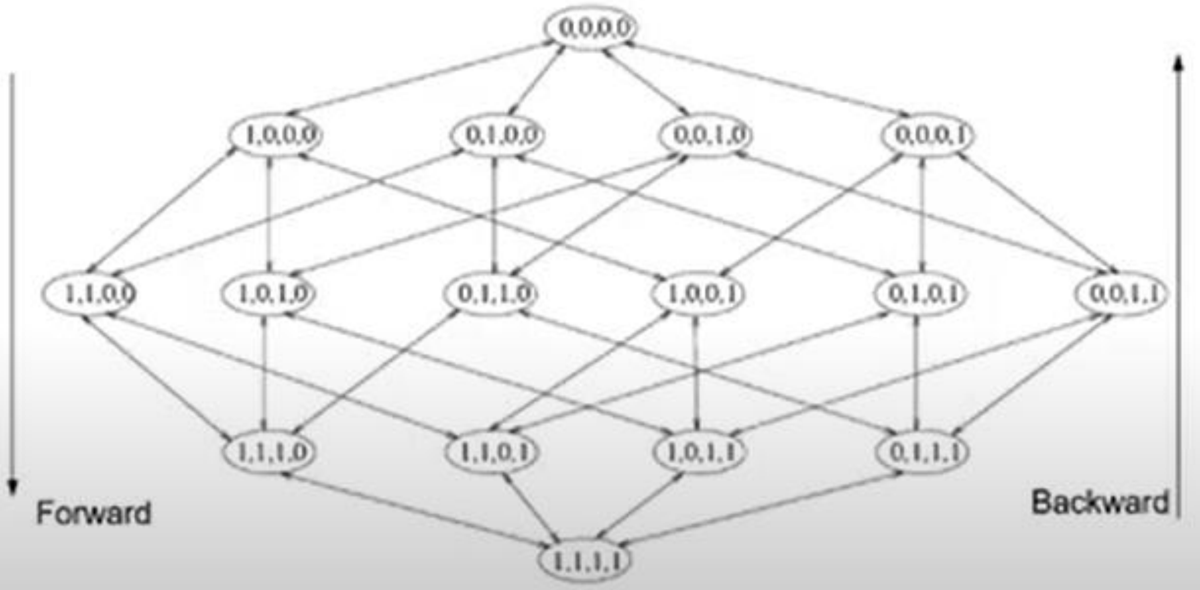
- Feature reduction
  - All original features are used
  - The transformed features are linear combinations of the original features
- Feature selection
  - Only a subset of the original features are selected

Basics

- Definitions of subset optimality
- Perspectives of feature selection
  - Subset search and feature ranking
  - Feature/subset evaluation measures
  - Models: filter vs. wrapper
  - Results validation and evaluation

A Subset Search Problem

- An example of search space (*Kohavi & John 1997*)



Feature Ranking

- Weighting and ranking individual features
- Selecting top-ranked ones for feature selection
- Advantages
  - Efficient: $O(N)$ in terms of dimensionality N
  - Easy to implement
- Disadvantages
  - Hard to determine the threshold
  - Unable to consider correlation between features

Evaluation Measures for Ranking and Selecting Features

- The goodness of a feature/feature subset is dependent on measures
- Various measures
  - Information measures (Yu & Liu 2004, Jebara & Jaakkola 2000)
    - (Entropy, Information gain)
  - Distance measures (Robnik & Kononenko 03, Pudil & Novovicov 98)
  - Dependence measures (Hall 2000, Modrzejewski 1993)
  - Consistency measures (Almuallim & Dietterich 94, Dash & Liu 03)
  - Accuracy measures (Dash & Liu 2000, Kohavi&John 1997)

Consistency Measures

- Consistency measures
  - Trying to find a minimum number of features that separate classes as consistently as the full set can
  - An inconsistency is defined as two instances having the same feature values but different classes
    - E.g., one inconsistency is found between instances i4 and i8 if we just look at the first two columns of the data table

Illustrative Data Set

	Hair	Height	Weight	Lotion	Result
i_1	1	2	1	0	1
i_2	1	3	2	1	0
i_3	2	1	2	1	0
i_4	1	1	2	0	1
i_5	3	2	3	0	1
i_6	2	3	3	0	0
i_7	2	2	3	0	0
i_8	1	1	1	1	0

Sunburn data

	Result (Sunburn)	
	No	Yes
P(Result)	5/8	3/8
P(Hair=1|Result)	2/5	2/3
P(Hair=2|Result)	3/5	0
P(Hair=3|Result)	0	1/3
P(Height=1|Result)	2/5	1/3
P(Height=2|Result)	1/5	2/3
P(Height=3|Result)	2/5	0
P(Weight=1|Result)	1/5	1/3
P(Weight=2|Result)	2/5	1/3
P(Weight=3|Result)	2/5	1/3
P(Lotion=0|Result)	2/5	3/3
P(Lotion=1|Result)	3/5	0

Priors and class conditional probabilities

Accuracy Measures

- Using classification accuracy of a classifier as an evaluation measure
- Factors constraining the choice of measures
  - Classifier being used
  - The speed of building the classifier
- Compared with previous measures
  - Directly aimed to improve accuracy
  - Biased toward the classifier being used
  - More time consuming

Models of Feature Selection

- Filter model
  - Separating feature selection from classifier learning
  - Relying on general characteristics of data (*information, distance, dependence, consistency*)
  - No bias toward any learning algorithm, fast
- Wrapper model
  - Relying on a predetermined classification algorithm
  - Using predictive accuracy as goodness measure
  - High accuracy, computationally expensive

Weka – Iris flower dataset

- 3 types (classes)
- 4 features /variables
- 150 instances/cases

Iris Versicolor Iris Setosa Iris Virginica

Iris dataset - features

- Sepal length
- Sepal width

- Petal length
- Petal width