Aalto University
School of Electrical Engineering

CS-C3240 – Machine Learning D

Round 3: From features to classification

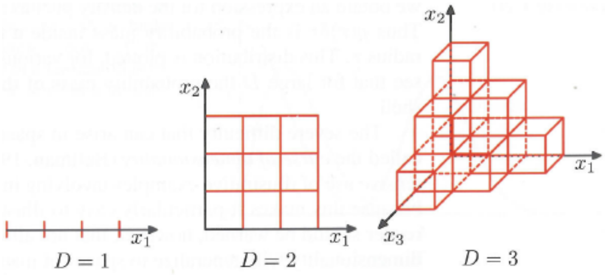
Stephan Sigg

Department of Communications and Networking
Aalto University, School of Electrical Engineering
stephan.sigg@aalto.fi

Version 2.3, February 12, 2022

Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

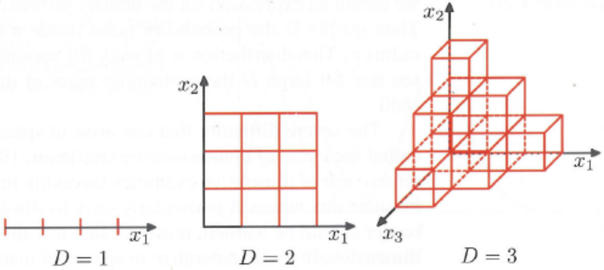


Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Curse of dimensionality

Too sparse samples across regions to estimate a distribution in that space
(Problematic for methods that require statistical significance)



Issues related to high dimensional input data

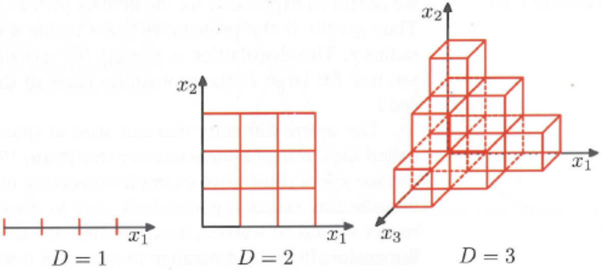
Exponential growth Volume of the space grows exponentially with dimension

Curse of dimensionality

Too sparse samples across regions to estimate a distribution in that space
(Problematic for methods that require statistical significance)

Hughes (peaking) phenomenon

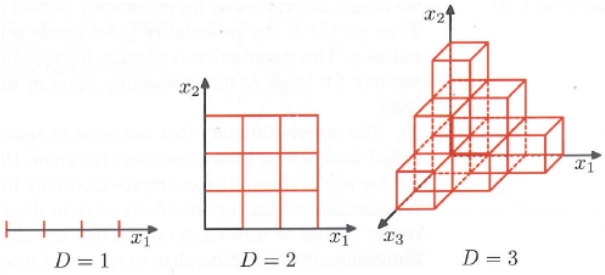
Predictive power of classifier first increases with dimension, then decreases



Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces



Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces

Example – Volume of a sphere

Consider a sphere of radius $r = 1$ in a D-dimensional space

Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces

Example – Volume of a sphere

Consider a sphere of radius $r = 1$ in a D-dimensional space

Fraction of the volume between radius $r = 1$ and $r' = 1 - \varepsilon$?

Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces

Example – Volume of a sphere

Consider a sphere of radius $r = 1$ in a D-dimensional space

Fraction of the volume between radius $r = 1$ and $r' = 1 - \varepsilon$?

Volume of shpere with radius r:

$V_D(r) = \delta_D r^D$ for appropriate δ_D

Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces

Example – Volume of a sphere

Consider a sphere of radius $r = 1$ in a D-dimensional space

Fraction of the volume between radius $r = 1$ and $r' = 1 - \varepsilon$?

Volume of shpere with radius r:

$$V_D(r) = \delta_D r^D \qquad \text{for appropriate } \delta_D$$

Given by

$$\frac{V_D(1) - V_D(1-\varepsilon)}{V_D(1)} = 1 - (1-\varepsilon)^D$$

Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces

Example – Volume of a sphere

Consider a sphere of radius $r = 1$ in a D-dimensional space

Fraction of the volume between radius $r = 1$ and $r' = 1 - \varepsilon$?

Volume of shpere with radius r:

$$V_D(r) = \delta_D r^D$$

for appropriate δ_D

Given by

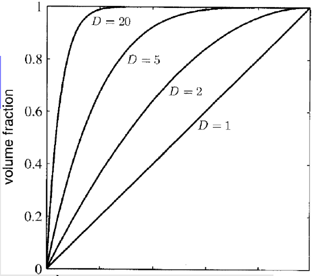
$$\frac{V_D(1) - V_D(1-\varepsilon)}{V_D(1)} = 1 - (1-\varepsilon)^D$$

For large D, this fraction tends to 1

In high dimensions, most of the volume of a sphere concentrates near the surface

Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension
Counter-intuitive properties in higher dimensional spaces



Example – Volume of a sphere

Consider a sphere of radius $r = 1$ in a D-dimensional space

Fraction of the volume between radius $r = 1$ and $r' = 1 - \varepsilon$?

Volume of shpere with radius r:

$$V_D(r) = \delta_D r^D$$ for appropriate δ_D

Given by

$$\frac{V_D(1) - V_D(1-\varepsilon)}{V_D(1)} = 1 - (1-\varepsilon)^D$$

For large D, this fraction tends to 1

In high dimensions, most of the volume of a sphere concentrates near the surface

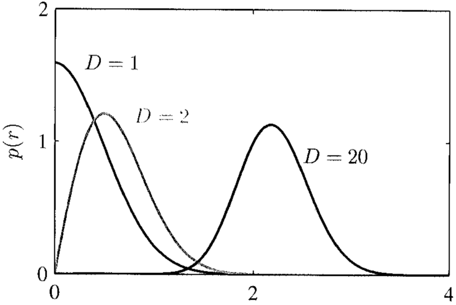
Issues related to high dimensional input data

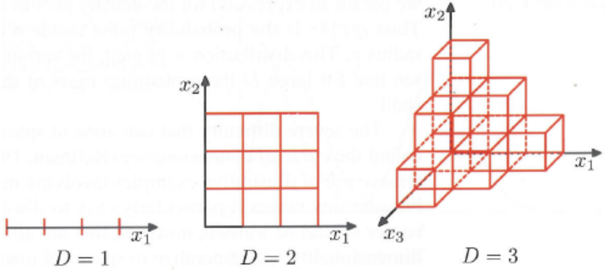
Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces

Example – Gaussian distribution

Probability mass concentrated in a thin shell
(here plotted as distance from the origin in a polar coordinate system)





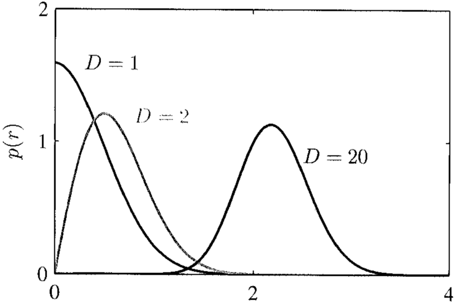
Issues related to high dimensional input data

Exponential growth Volume of the space grows exponentially with dimension

Counter-intuitive properties in higher dimensional spaces

Example – Gaussian distribution

Probability mass concentrated in a thin shell
(here plotted as distance from the origin in a polar coordinate system)



Curse of Dimensionality

Mechanisms to efficiently reduce dimensions or classifiers that respect properties of high-dimensional spaces required.

Questions?

Stephan Sigg
stephan.sigg@aalto.fi

Si Zuo
si.zuo@aalto.fi

Literature

- C.M. Bishop: Pattern recognition and machine learning, Springer, 2007.
- R.O. Duda, P.E. Hart, D.G. Stork: Pattern Classification, Wiley, 2001.