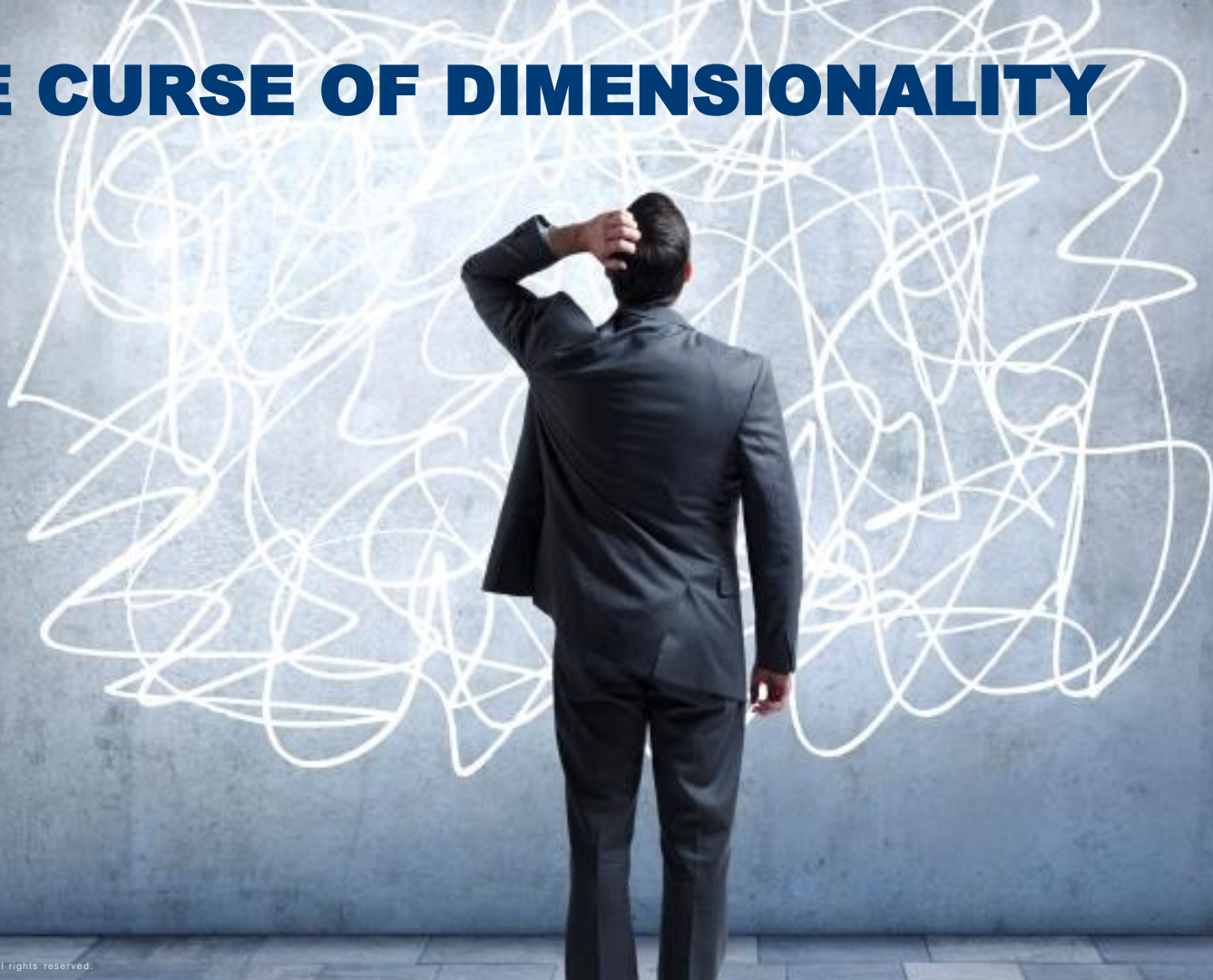


**VARIABELSELEKTERING –VIKTIGARE ÄN VAD DU TROR!**  
"THE CURSE OF DIMENSIONALITY" ÄR ETT BEGREPP INOM MODELLERING - HUR UNDVIKER VI  
DETTA?



# THE CURSE OF DIMENSIONALITY

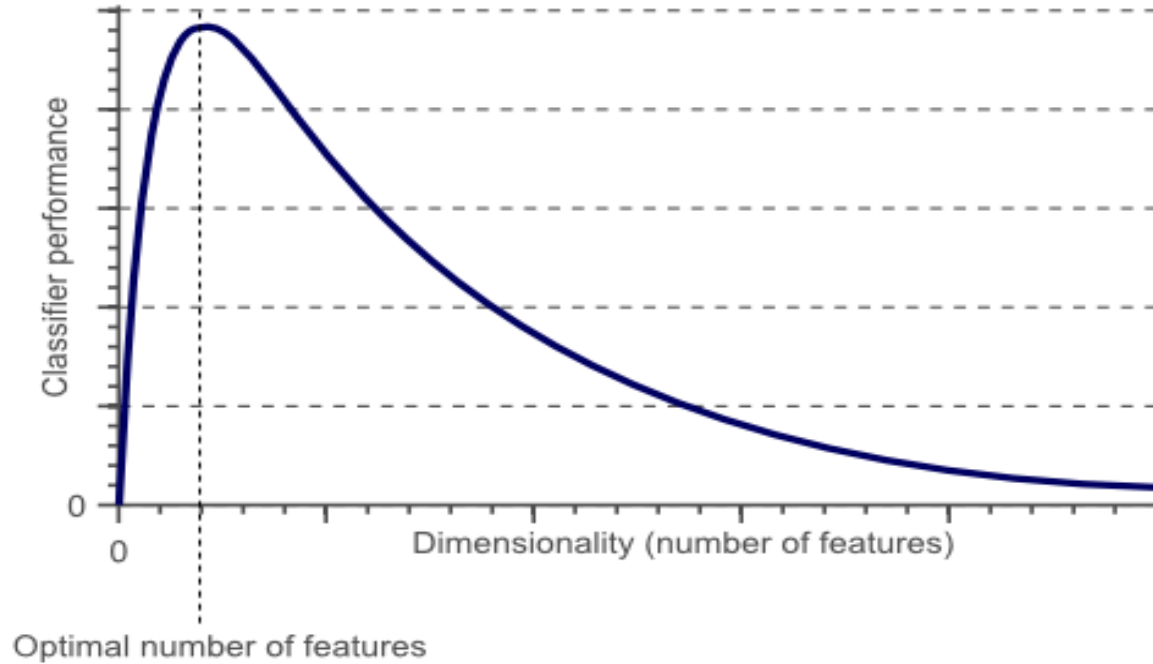




# THE CURSE OF DIMENSIONALITY



# THE CURSE OF DIMENSIONALITY



# BOTEMEDEL?

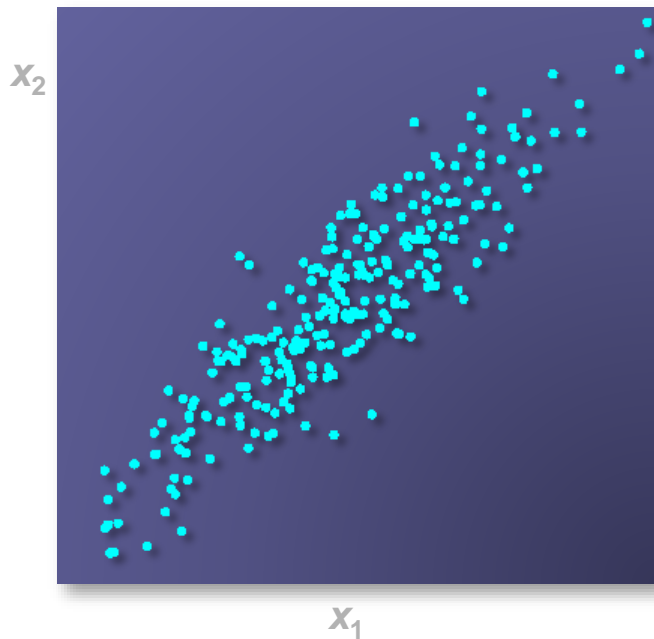




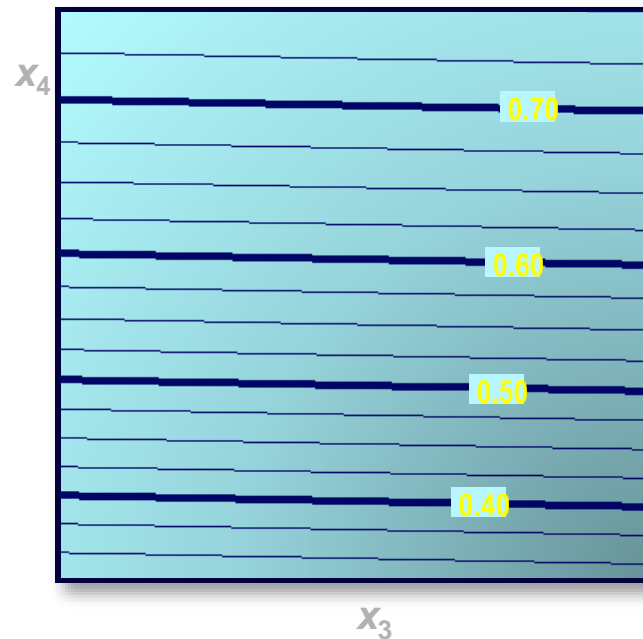
# THE CURSE OF DIMENSIONALITY

## INPUT STRATEGIES

### Redundancy



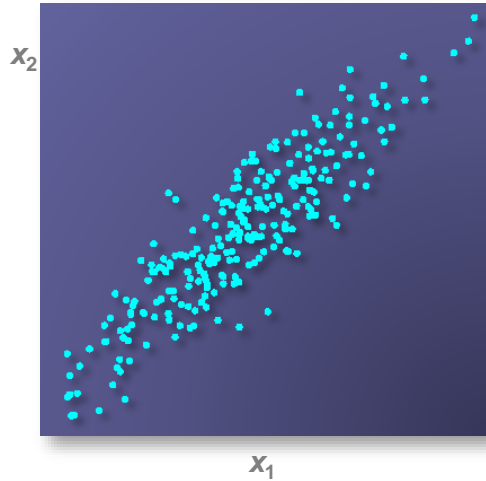
### Irrelevancy



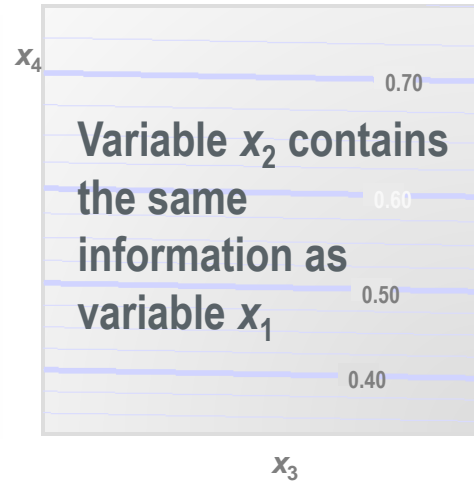
# THE CURSE OF DIMENSIONALITY

## INPUT REDUCTION STRATEGIES

Redundancy



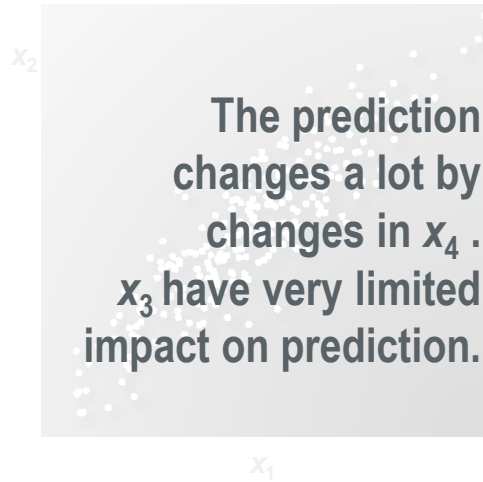
Irrelevancy



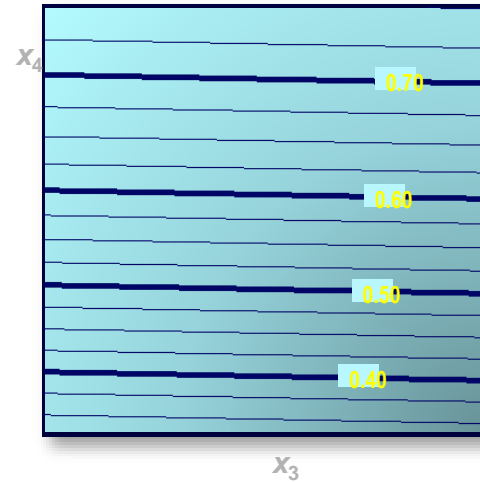
...

## INPUT REDUCTION STRATEGIES

Redundancy



Irrelevancy

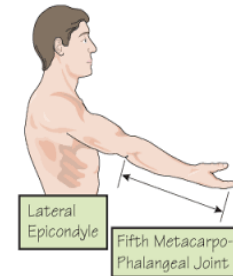
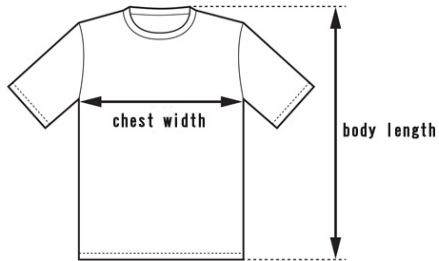
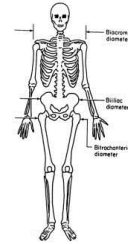
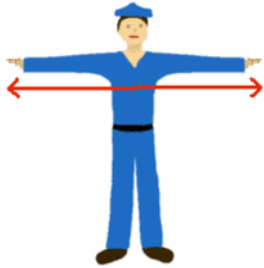




- Supervised(prediktionsförmåga-koppling till Målvariabel)
  - Modeller med en variabel mot target
  - Informationsvärde/GINI, etc.
  - Affärskunskap
- Unsupervised(samband mellan inputs)
  - Variabelklustring
  - PCA
  - Klusteranalys

# VARIABEL SELEKTERING

## VARIABELKLUSTER(PROC VARCLUS) EXEMPEL



# VARIABEL SELEKTERING

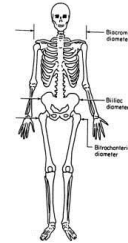
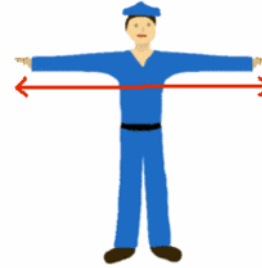
## VARIABELKLUSTER EXEMPEL

|          |      |      |      |      |      |      |      |      |
|----------|------|------|------|------|------|------|------|------|
| Height   | 1.0  | .846 | .805 | .859 | .473 | .398 | .301 | .382 |
| ArmSpan  | .846 | 1.0  | .881 | .826 | .376 | .326 | .277 | .415 |
| Forearm  | .805 | .881 | 1.0  | .801 | .380 | .319 | .237 | .345 |
| LowerLeg | .859 | .826 | .801 | 1.0  | .436 | .329 | .327 | .365 |
| Weight   | .473 | .376 | .380 | .436 | 1.0  | .762 | .730 | .629 |
| BitDiam  | .398 | .326 | .319 | .329 | .762 | 1.0  | .583 | .577 |
| Girth    | .301 | .277 | .237 | .327 | .730 | .583 | 1.0  | .539 |
| Width    | .382 | .415 | .345 | .365 | .629 | .577 | .539 | 1.0  |

# VARIABEL SELEKTERING

## VARIABELKLUSTER EXEMPEL

| 2 Clusters |          | R-squared with |                 | 1-R**2<br>Ratio | Variable<br>Label       |
|------------|----------|----------------|-----------------|-----------------|-------------------------|
| Cluster    | Variable | Own<br>Cluster | Next<br>Closest |                 |                         |
| Cluster 1  | ArmSpan  | 0.9002         | 0.1658          | 0.1196          | Arm Span                |
|            | Forearm  | 0.8661         | 0.1413          | 0.1560          | Length of Forearm       |
|            | LowerLeg | 0.8652         | 0.1829          | 0.1650          | Length of Lower Leg     |
|            | Height   | 0.8777         | 0.2088          | 0.1545          |                         |
| Cluster 2  | BitDiam  | 0.7386         | 0.1341          | 0.3019          | Bitrochanteric Diameter |
|            | Girth    | 0.6981         | 0.0929          | 0.3328          | Chest Girth             |
|            | Width    | 0.6329         | 0.1619          | 0.4380          | Chest Width             |
|            | Weight   | 0.8477         | 0.1974          | 0.1898          |                         |

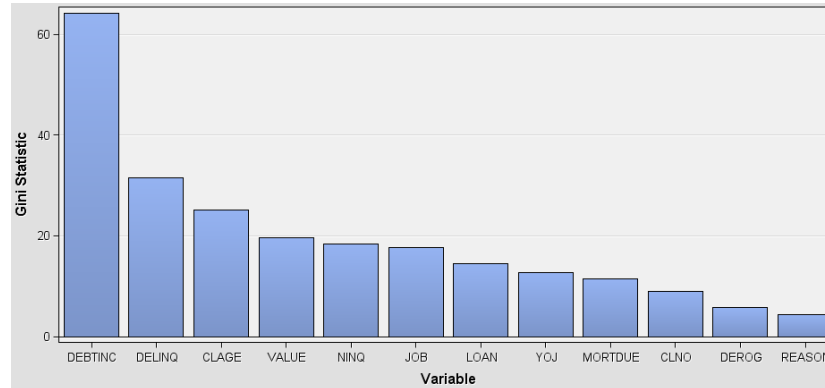




# VARIABEL SELEKTERING

## BINNING

| Variable | Gini Statistic | Level for Interactive | Calculated Role | New Role | Level    | Label | Gini Ordering |
|----------|----------------|-----------------------|-----------------|----------|----------|-------|---------------|
| DEBTINC  | 64.261         | INTERVAL              | Input           | Default  | INTERVAL |       | 1             |
| DELINQ   | 31.588         | INTERVAL              | Input           | Default  | INTERVAL |       | 2             |
| CLAGE    | 25.095         | INTERVAL              | Input           | Default  | INTERVAL |       | 3             |
| VALUE    | 19.702         | INTERVAL              | Rejected        | Default  | INTERVAL |       | 4             |
| NINQ     | 18.346         | INTERVAL              | Rejected        | Default  | INTERVAL |       | 5             |
| JOB      | 17.626         | NOMINAL               | Rejected        | Default  | NOMINAL  |       | 6             |
| LOAN     | 14.435         | INTERVAL              | Rejected        | Default  | INTERVAL |       | 7             |
| YOJ      | 12.633         | INTERVAL              | Rejected        | Default  | INTERVAL |       | 8             |
| MORTDUE  | 11.436         | INTERVAL              | Rejected        | Default  | INTERVAL |       | 9             |
| CLNO     | 9.049          | INTERVAL              | Rejected        | Default  | INTERVAL |       | 10            |
| DEROG    | 5.699          | INTERVAL              | Rejected        | Default  | INTERVAL |       | 11            |
| REASON   | 4.311          | NOMINAL               | Rejected        | Default  | NOMINAL  |       | 12            |



**TACK FÖR MIG**  
**MATHIAS.LANNER@SAS.COM**



**THE  
POWER  
TO KNOW.**