

Bag of Visual Words Model

Albert Chung

Department of Computer Science

University of Exeter

Spring Term 2023

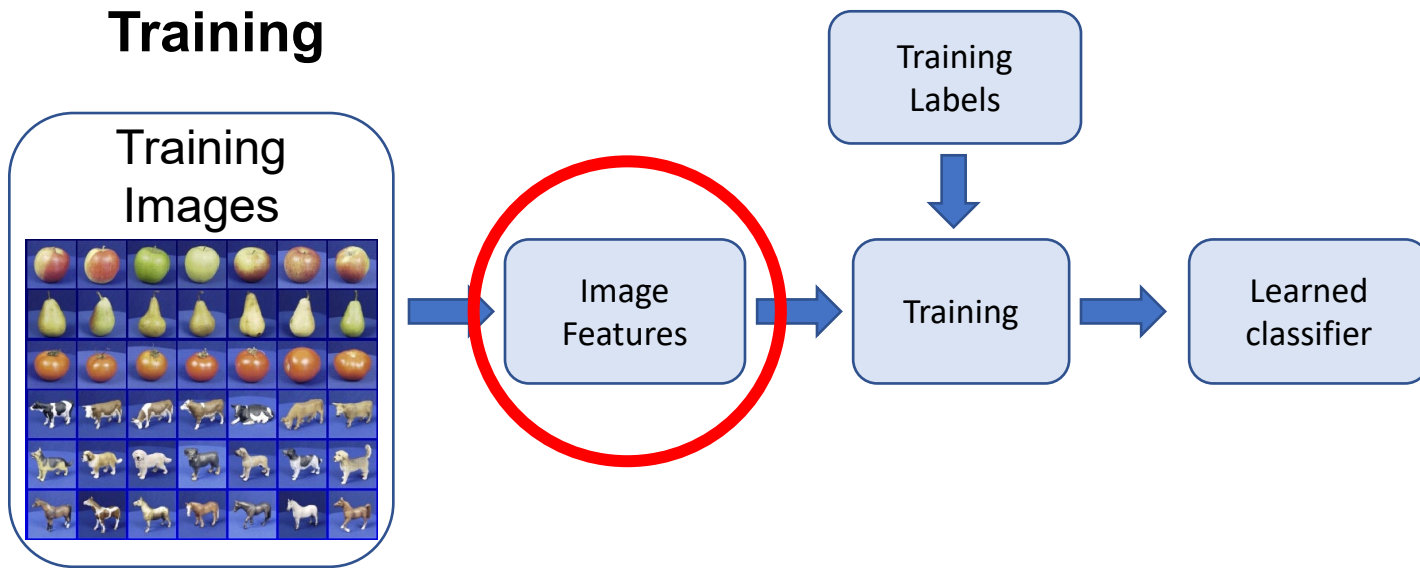
References

- R. Szeliski, “Computer Vision: Algorithms and Applications” (2nd Edition):
 - Image Classification: Chapter 6.2
 - Large-scale Matching and Retrieval: Chapter 7.1.4
- C. M. Bishop, “Pattern Recognition and Machine Learning”:
 - Maximum Margin Classifiers: Chapter 7.1

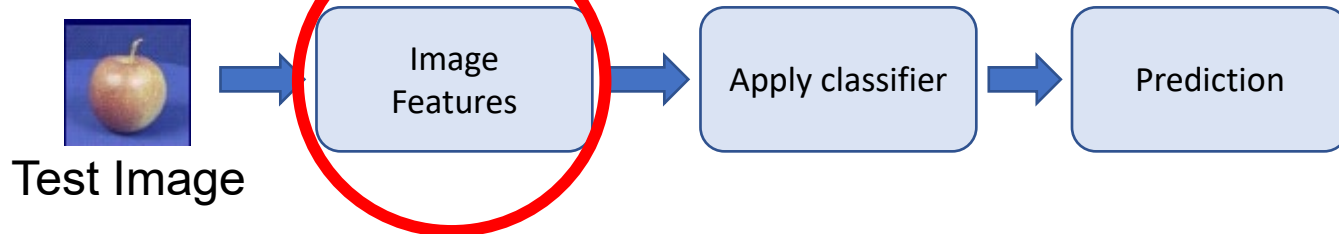
Image classification



Training



Testing



Motivation: Bags of words (BoW)

- Order-less document representation: frequencies of words from a dictionary



Motivation: Bags of words (BoW), text mining, and similar document search

- Order-less document representation: frequencies of words from a dictionary

1863-01-01: Emancipation Proclamation

Abraham Lincoln (1861-65)

abolish acknowledged aforesaid african afterwards aggregate agriculture appropriation armed army assurance boundary burdens carolina census circulation colonization commanders communicated **compensation** complaints consent **constitution** contemplated convention currency **debt** deem **deportation** designated district doubted economical elections **emancipation** europe evils expended **fact** faithfully freed **freedom** god gradual herewith hostilities illinois **improving** independence indian indispensable inhabit injurious insurgents intercourse **january** judicial june **labor** legislatures liberation lord loyal mountains navy negotiation norfolk obligation opinion permanently perpetuate preserving proclamation proposition prosperity **rebellion** restoration revenue rivers seceded senate separation september **slave** **slavery** spain suppression telegraph **territory** therein thereof treasury treaty tribes true vessels virginia **war** wise york

1997-02-04: State of the Union Address

Bill Clinton (1993-2001)

agreement **americorps** asia ban **bipartisan** bless campaign celebration chamber china classroom college commitment conflicts constitution convention corporations **crime** criminals deficit democratic diversity earn economy education embrace enact endanger endeavor enduring europe expand exports families farsighted forty freedom **fueled** fundamental funding gangs global god hospital illegal immigrants incentives **internet** invest korea lifetime Locke love math **medicaid** **medicare** mexico nato northern nuclear obligation oklahoma partnership pension **polluters** prosperity renew republican restore reward richard rivers rolls russia safer science scientists sector **strength** students succeed **tax** teachers teaching teen **tejeda** terrorists threats toxic tuition tutors unfinished values violence violent **war** washington weapons welfare zarfos

2001-02-27: State of the Union Address

George W. Bush (2001-)

abusive administrator affordable agreement **bipartisan** bless capitol challenges chamber charities china civility classroom college commission **commitment** compassion confront congressman conservation crime debates **debt** dedicates deficit democrat deserve disabilities diseases earn economy education enforce equality expand failing fairness families foundation **freedom** **funding** global god ideals independence inflation **internet** invest invitation john **josefina** judged loved lowest math mayor **medicare** mentor neighborhoods nuclear overtime owe partners philadelphia pledge prescription presidential priority privilege prosperity punish recruit **refundable** regardless repaid restore reward science senate seniors steven strength students surpluses **targeted** **tax** teachers teaching terrorists threats transform trillion triple uncertainty values violence war washington weapons welfare

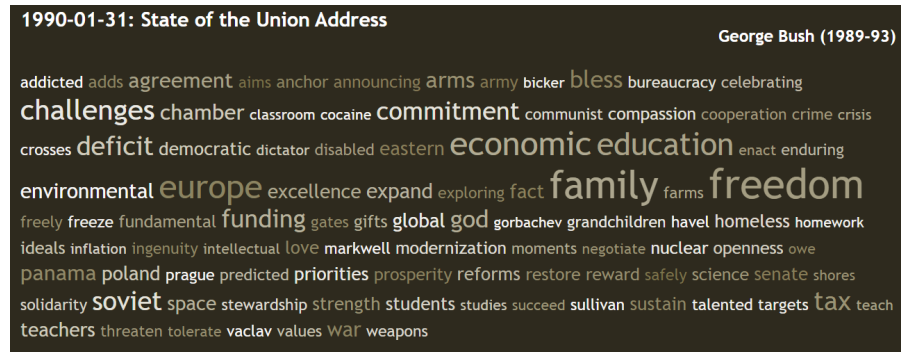
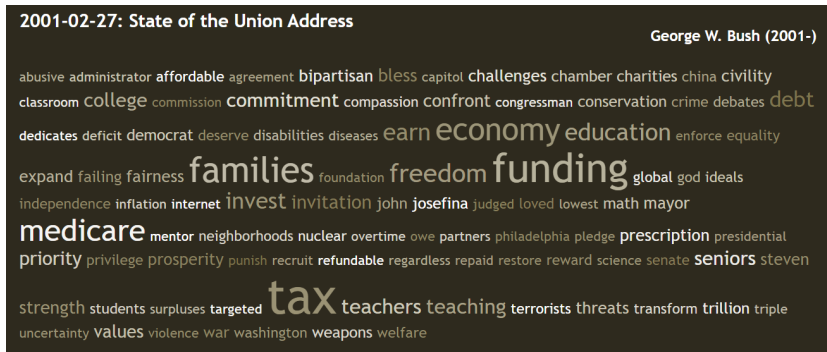
1990-01-31: State of the Union Address

George Bush (1989-93)

addicted adds agreement aims anchor announcing arms army bicker **bless** bureaucracy celebrating challenges chamber classroom cocaine **commitment** communist compassion cooperation crime crisis crosses deficit democratic dictator disabled eastern **economic** education enact enduring environmental europe excellence expand exploring fact family farms **freedom** freely freeze fundamental **funding** gates gifts global god gorbachev grandchildren havel homeless homework ideals inflation ingenuity intellectual love markwell modernization moments negotiate nuclear openness owe panama poland prague predicted priorities prosperity reforms restore reward safely science senate shores solidarity **soviet** space stewardship strength students studies succeed sullivan sustain talented targets **tax** teach teachers threaten tolerate vaclav values **war** weapons

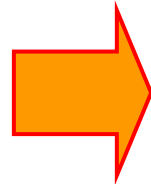
Motivation: Bags of words (BoW), text mining, and similar document search

- Order-less document representation: frequencies of words from a dictionary



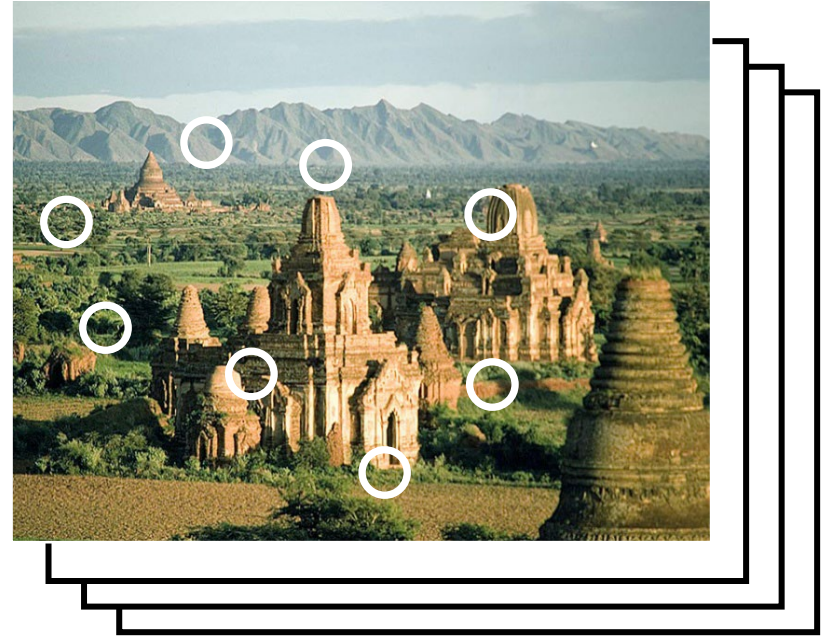
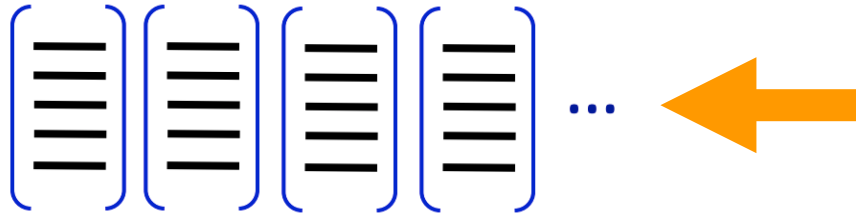
similar document

Bag of features / Bag of visual words

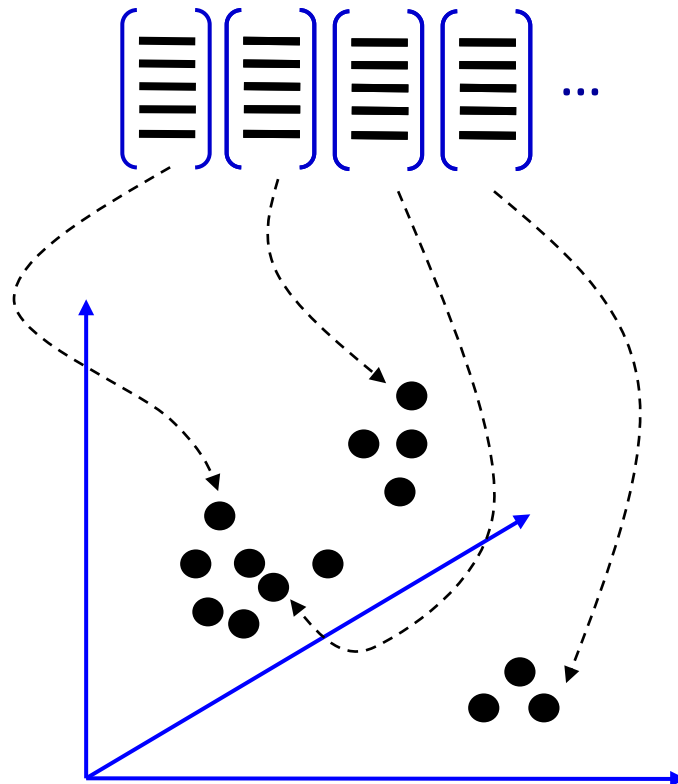


Feature Extraction

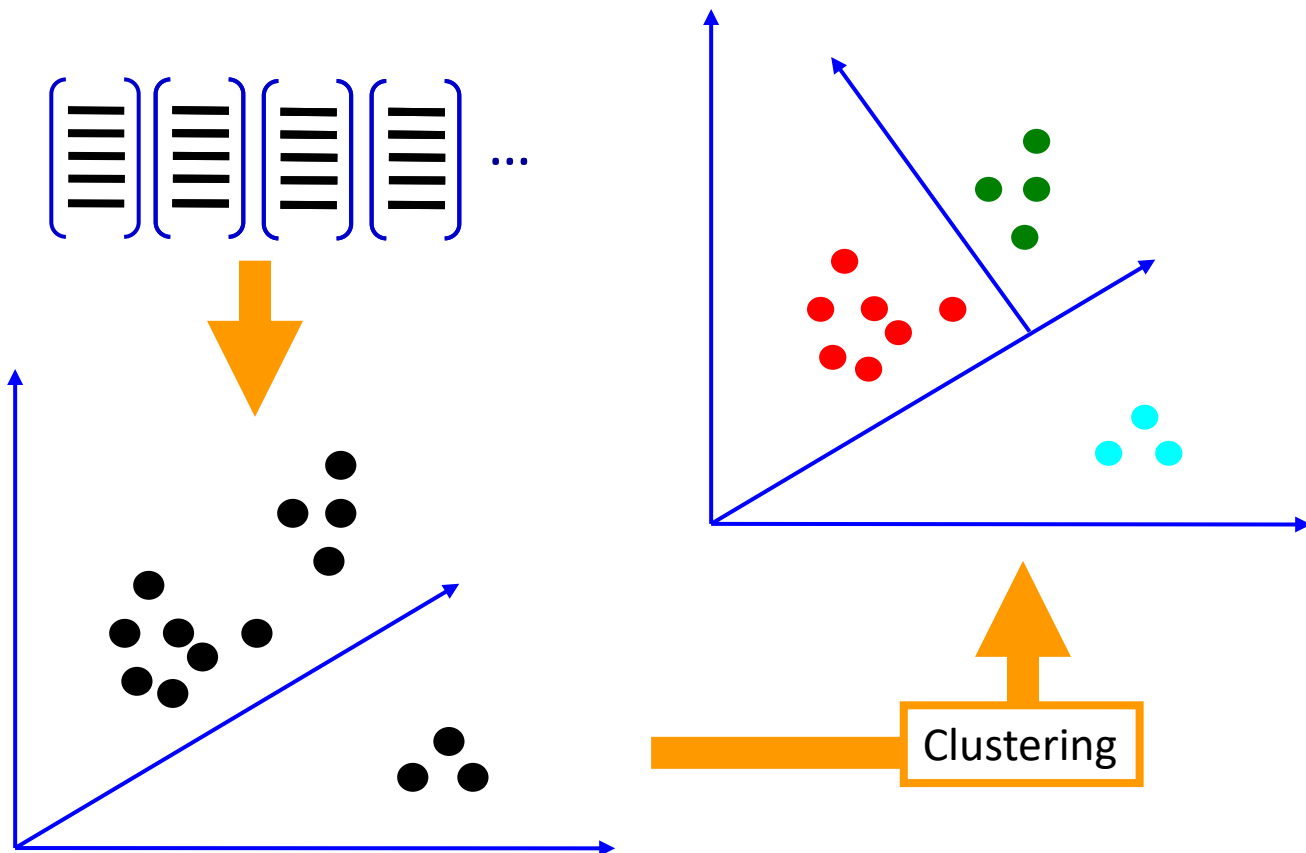
- Detecting features (e.g., corners) and computing descriptors from images
- Gathering descriptors from images in a collection



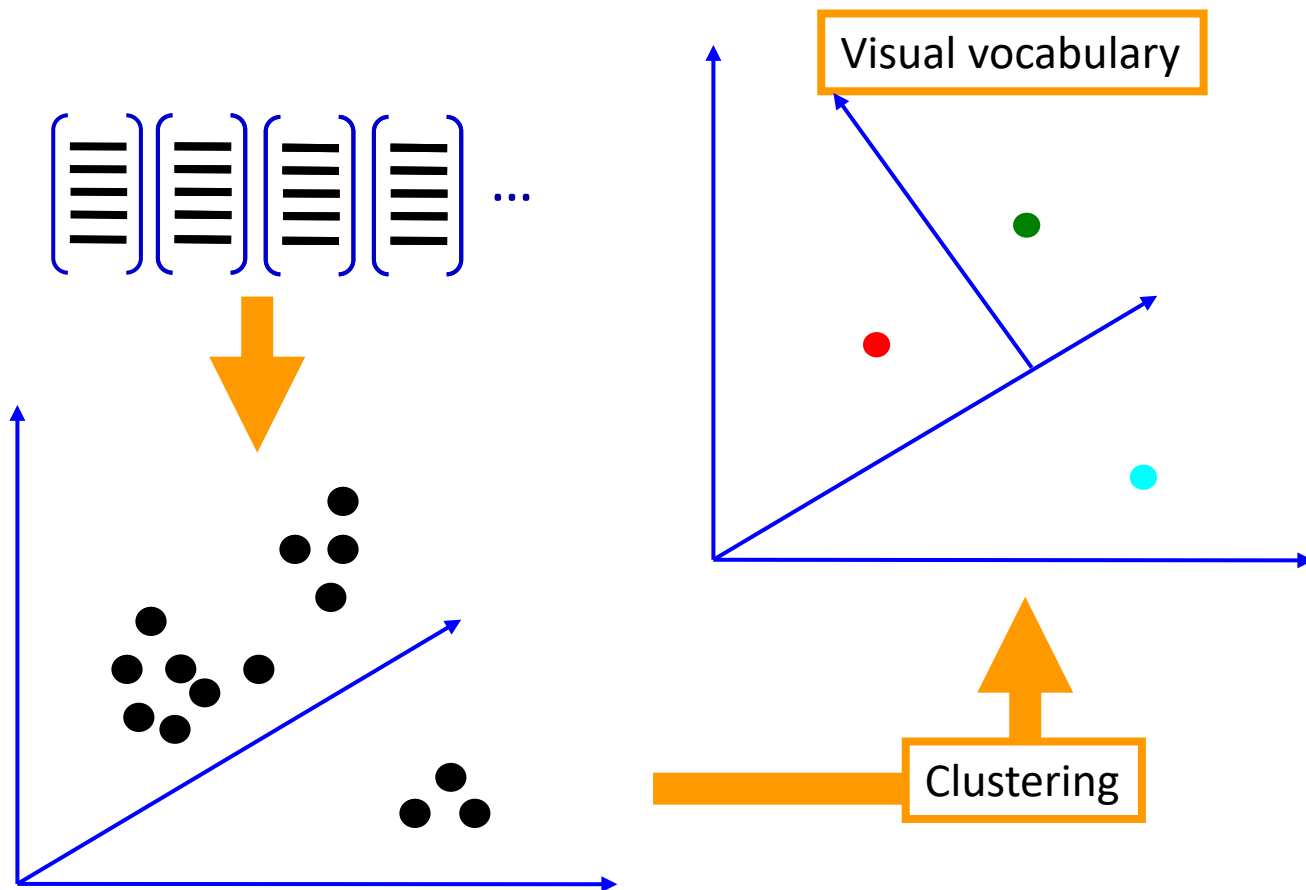
Extracted descriptors from the training set



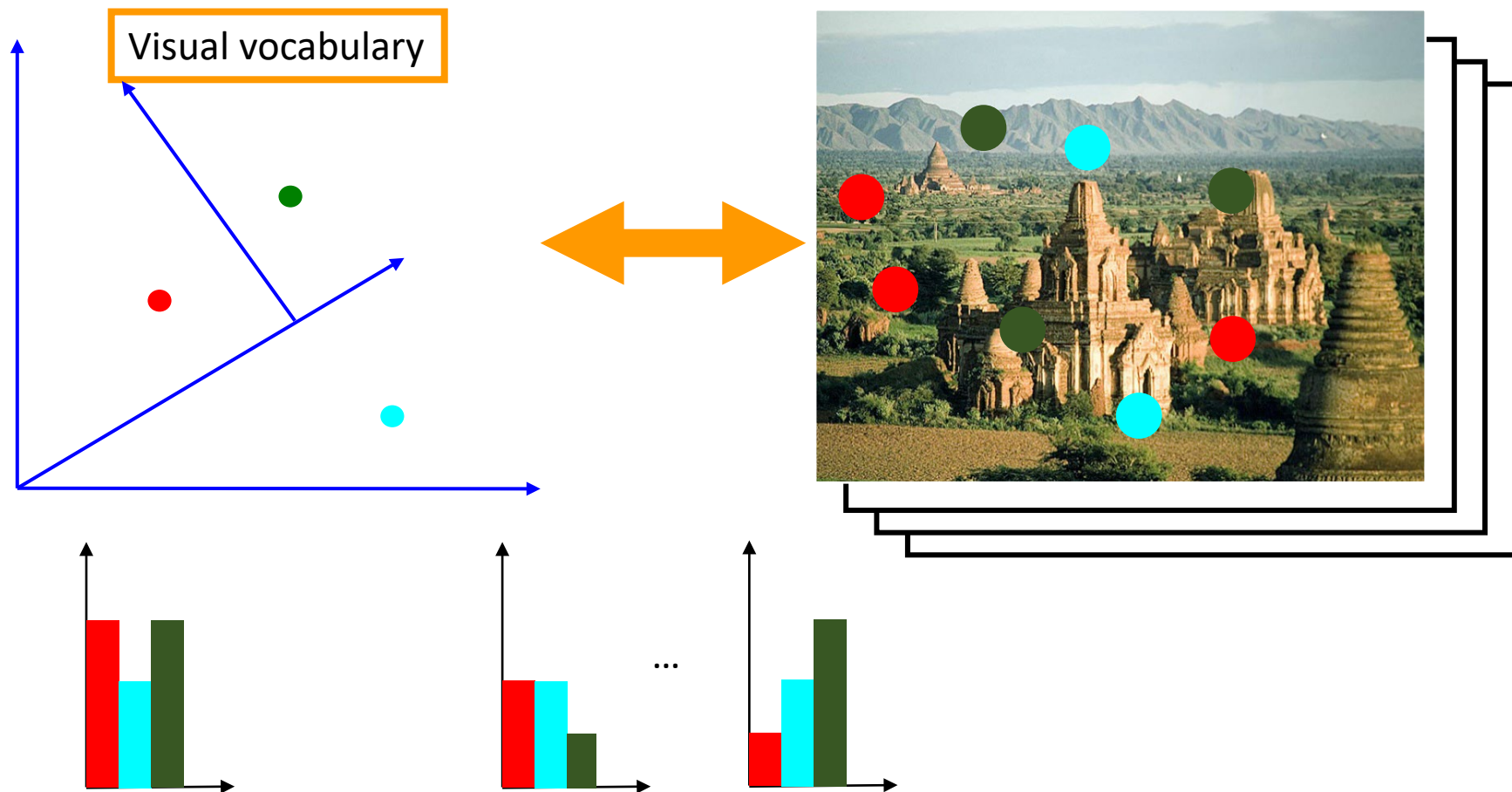
Learning the visual vocabulary



Learning the visual vocabulary or codebook



Feature Quantization

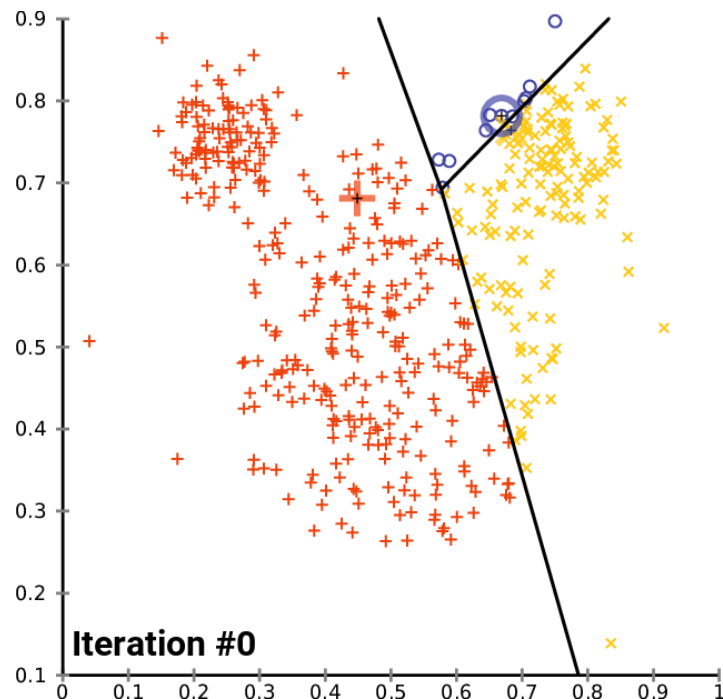


K-means clustering

- Want to minimize sum of squared Euclidean distances between features X_i and their nearest cluster centers m_k

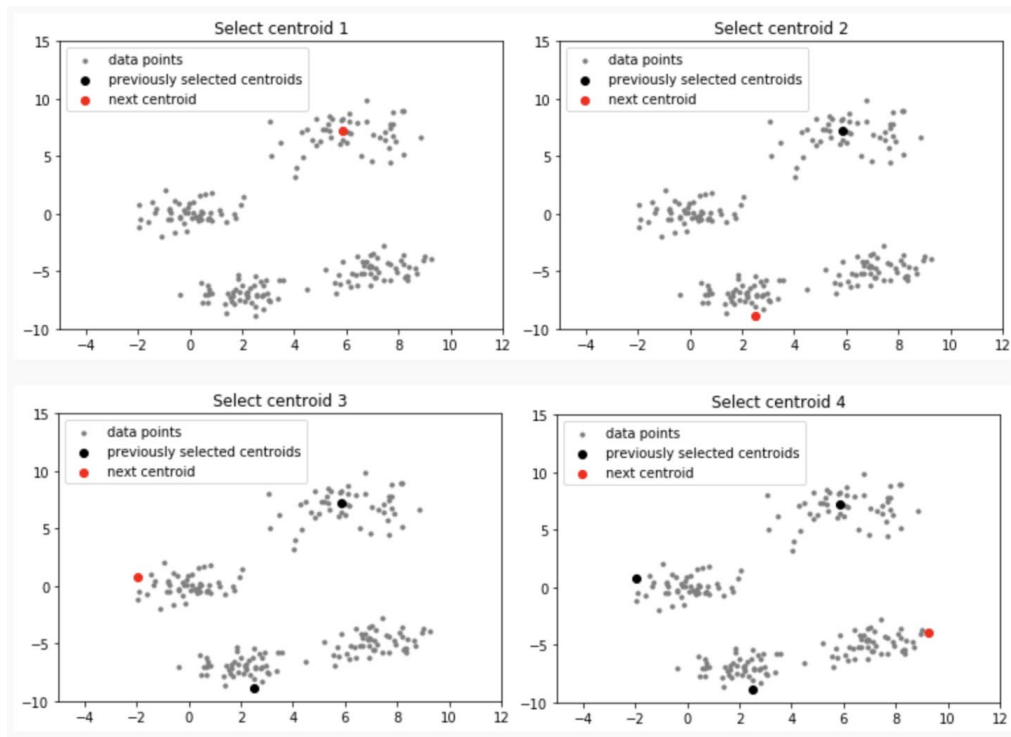
$$D(X, M) = \sum_k \sum_{i \in C_k} (X_i - m_k)^2$$

- Algorithm:
 - Randomly initialize K cluster centers
 - Iterate until convergence:
 - Assign each feature to the nearest center
 - Recompute each cluster center as the mean of all features assigned to it



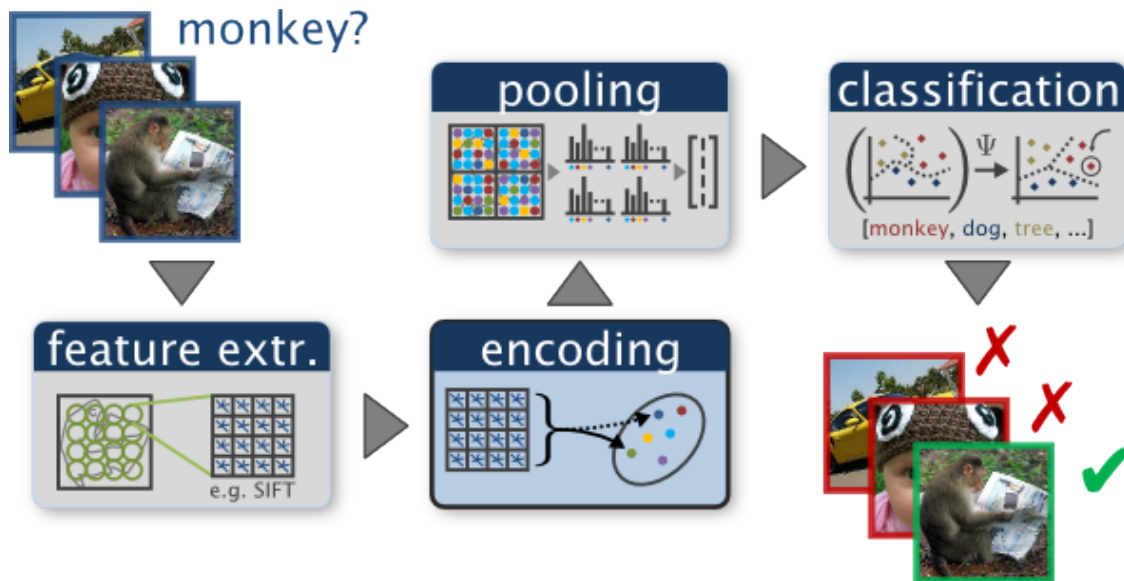
K-means++ clustering

- Algorithm:
 - [1] Randomly select the first cluster center
 - [2] Select the next cluster center which is spread out
 - Do [1] and [2] for k cluster centers



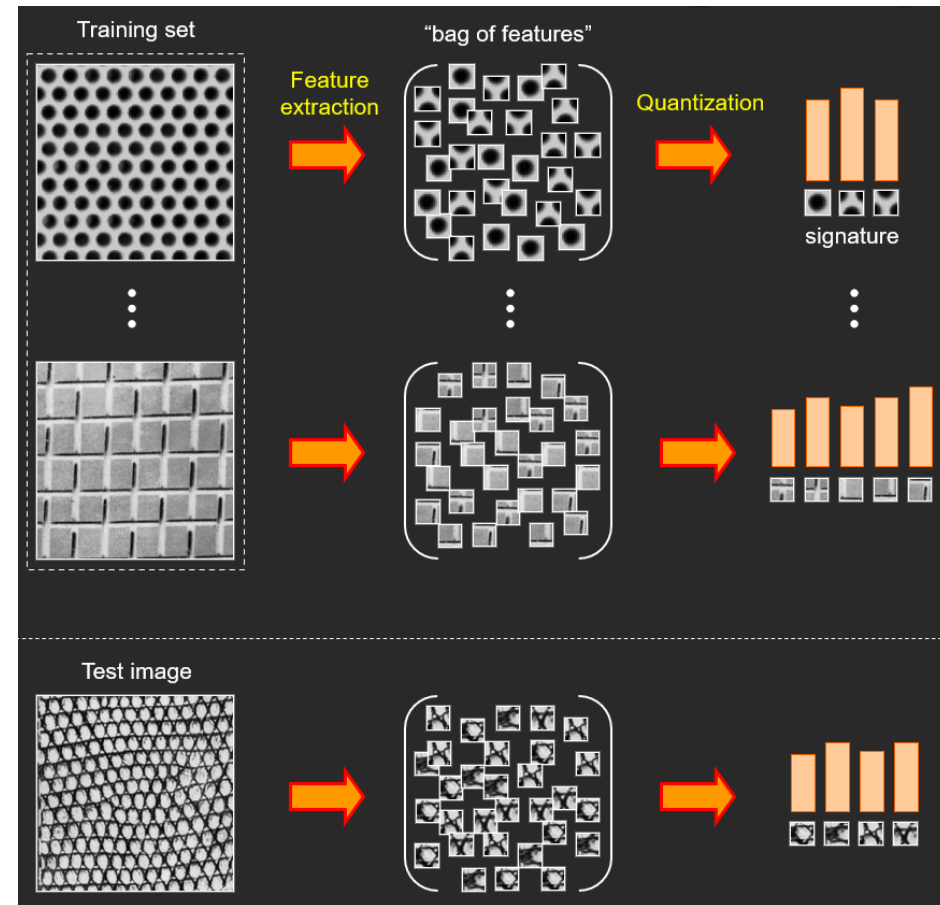
Bag of features: Outline

1. Extract local features
2. Learn visual vocabulary or cluster local features
3. Quantize or pool local features using visual vocabulary or codebook
4. Represent images by frequencies of “visual words”

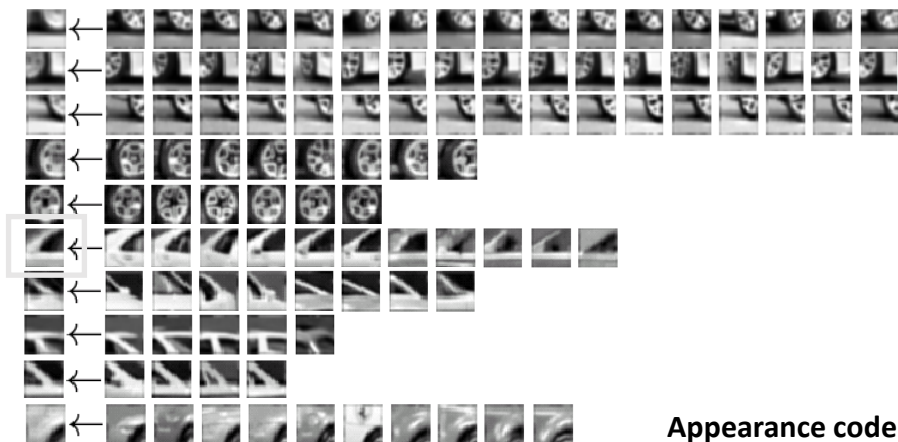
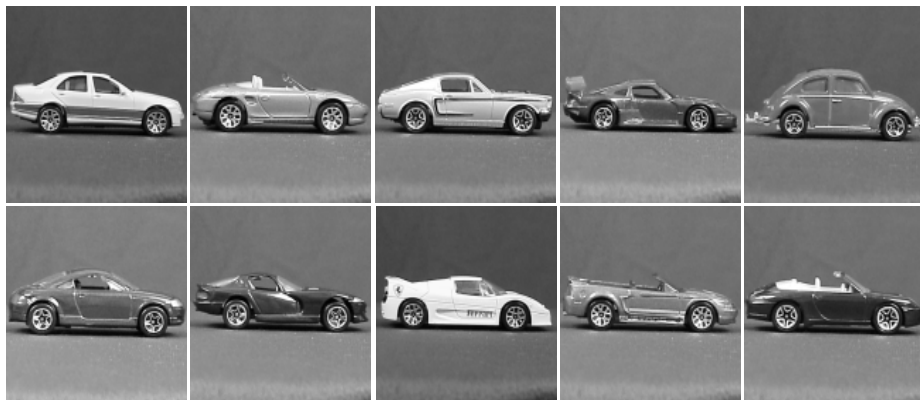


Bag of visual words

- The content can be inferred from the frequencies of words that happen in a document.
- Pixels are converted to visual words (a codebook/dictionary).
- The content can be inferred from the frequencies of visual words that appear in an image.



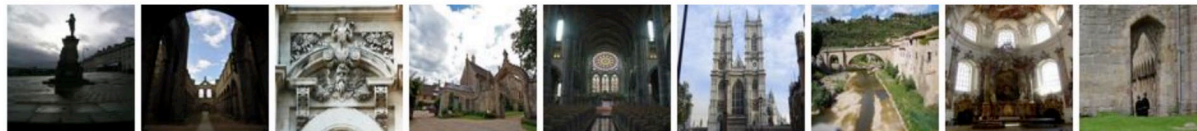
Recall: visual vocabularies



Appearance codebook

Necessity of spatial information

a_abbey(46368)



a_airfield(10910)



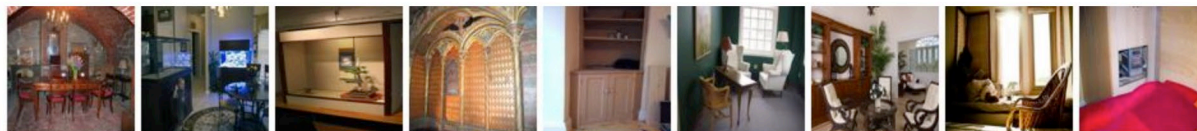
a_airplane_cabin(5152)



a_airport_terminal(16174)



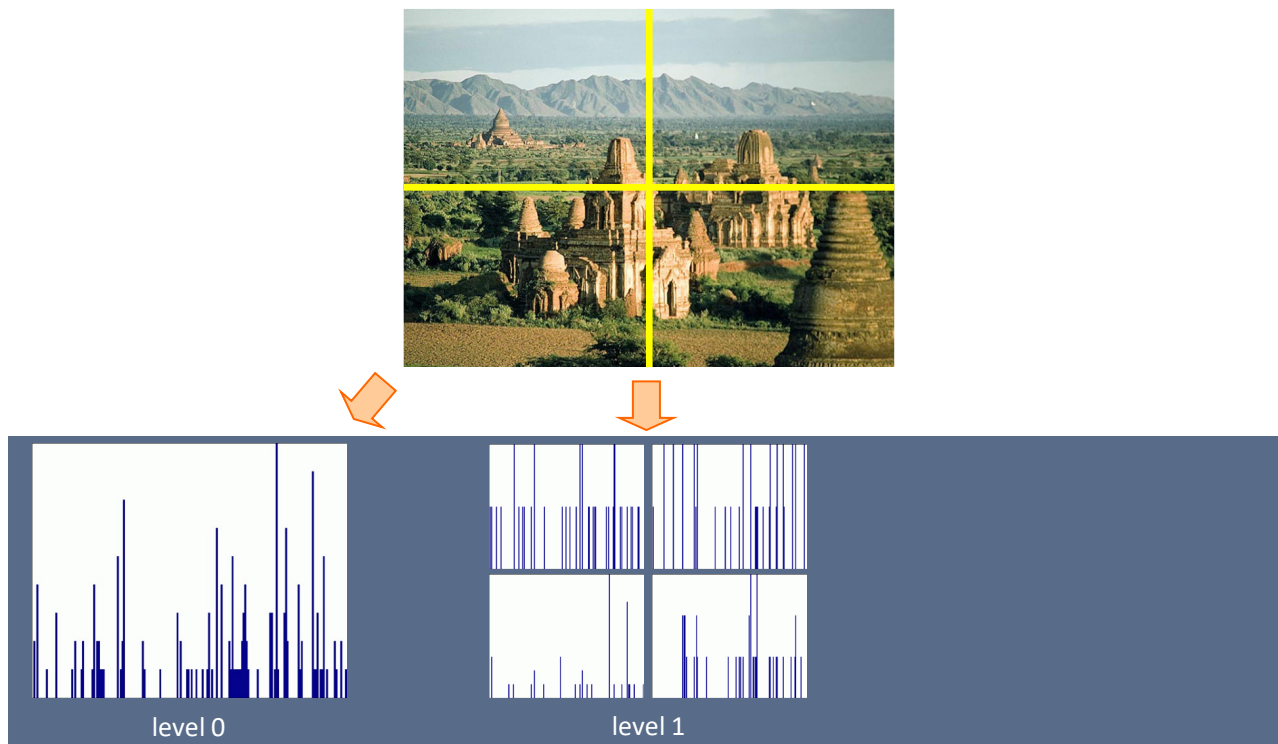
a_alcove(4966)



Spatial pyramids

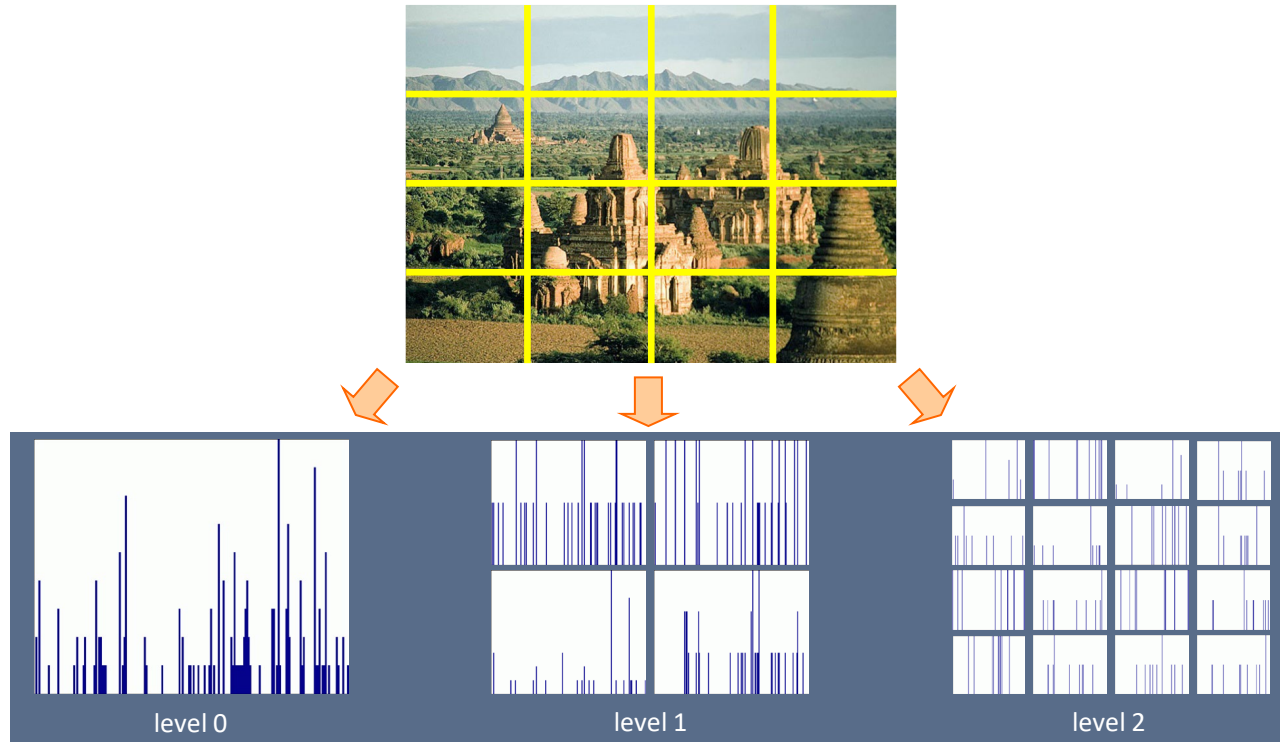


Spatial pyramids



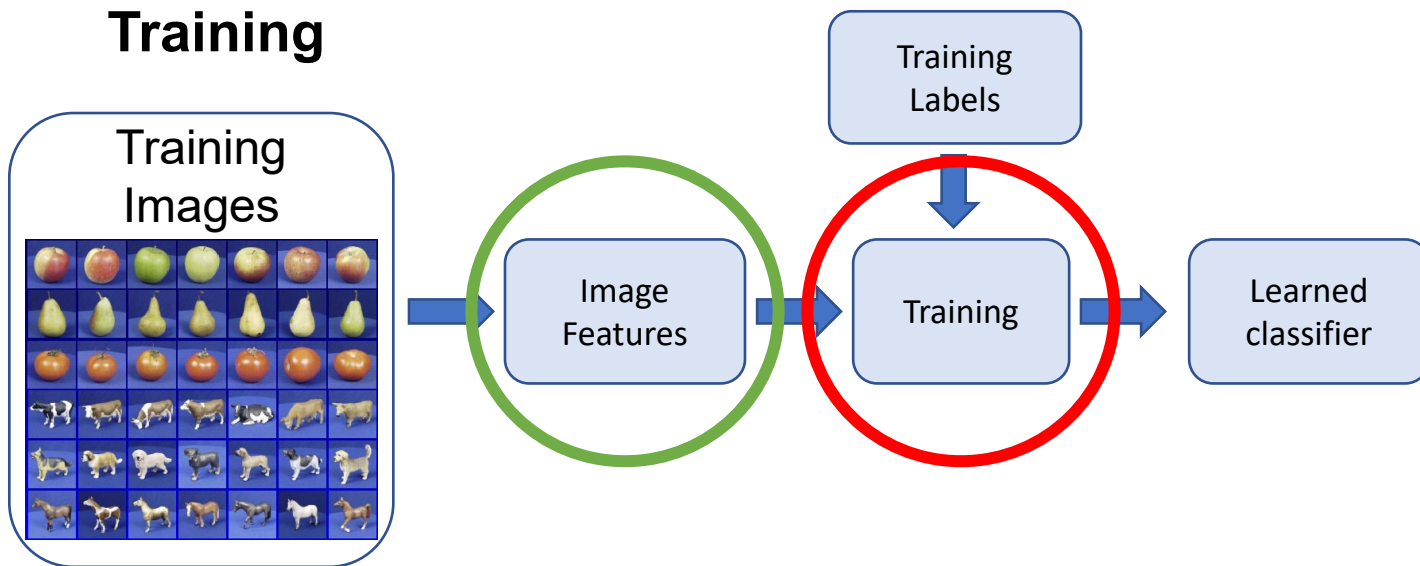
Lazebnik et al., Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR, 2006.

Spatial pyramids

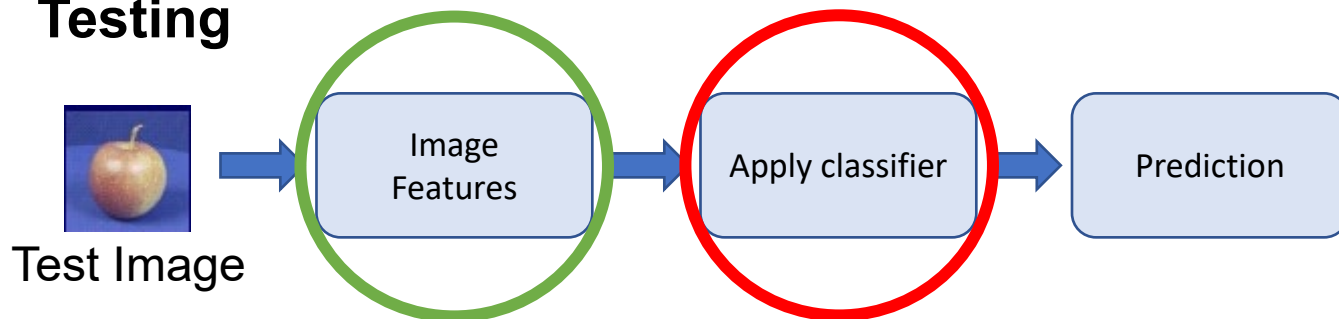


Lazebnik et al., Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR, 2006.

Training

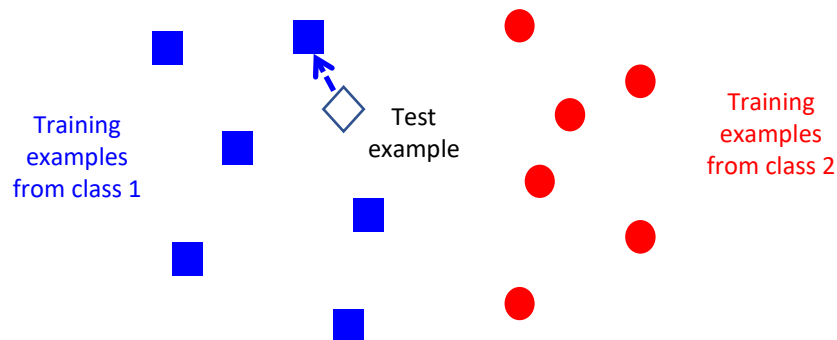


Testing



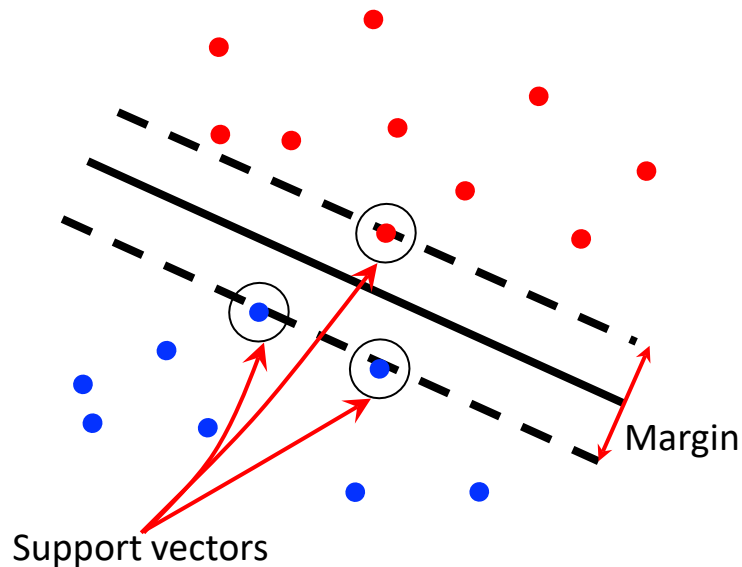
Nearest Neighbor Classifier

- $f(\mathbf{x})$ = label of the training example nearest to \mathbf{x}
- All we need is a distance or similarity function for our inputs



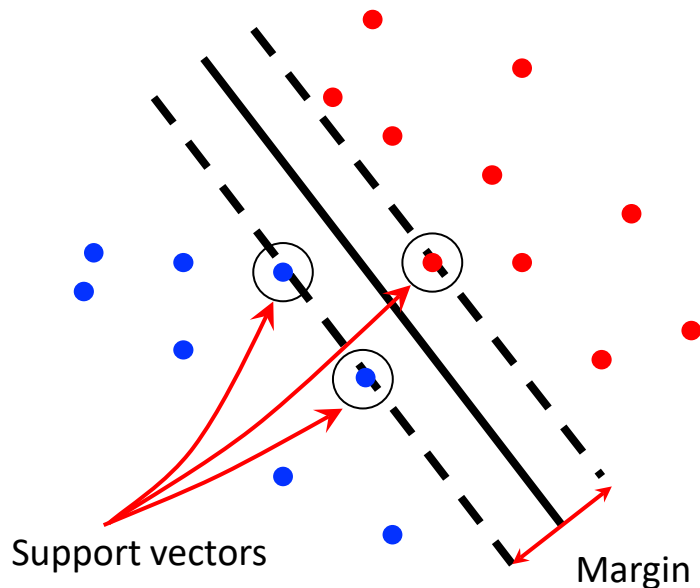
Large or Maximum Margin Classifier

- Find hyperplane that maximizes the margin between the positive and negative examples



Classifier: support vector machines

- Find hyperplane that maximizes the margin between the positive and negative examples



$$\mathbf{x}_i \text{ positive } (y_i = 1): \quad \mathbf{x}_i \cdot \mathbf{w} + b \geq 1$$

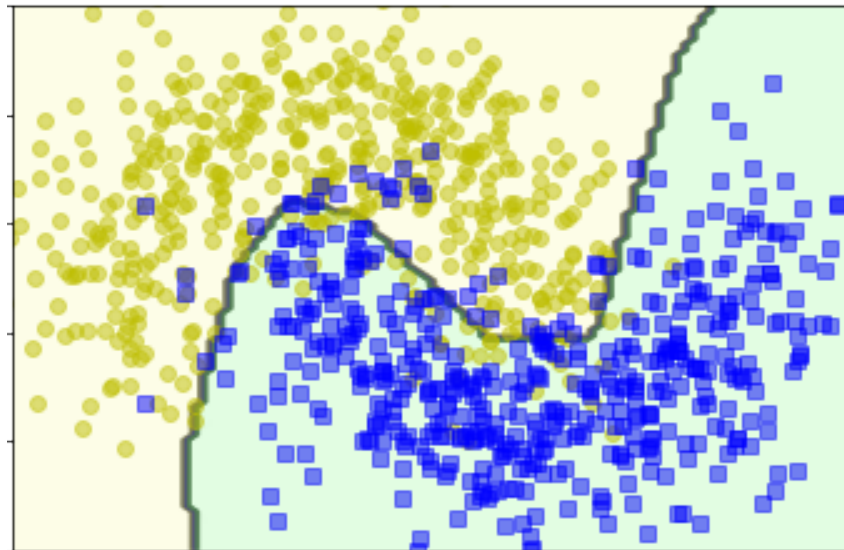
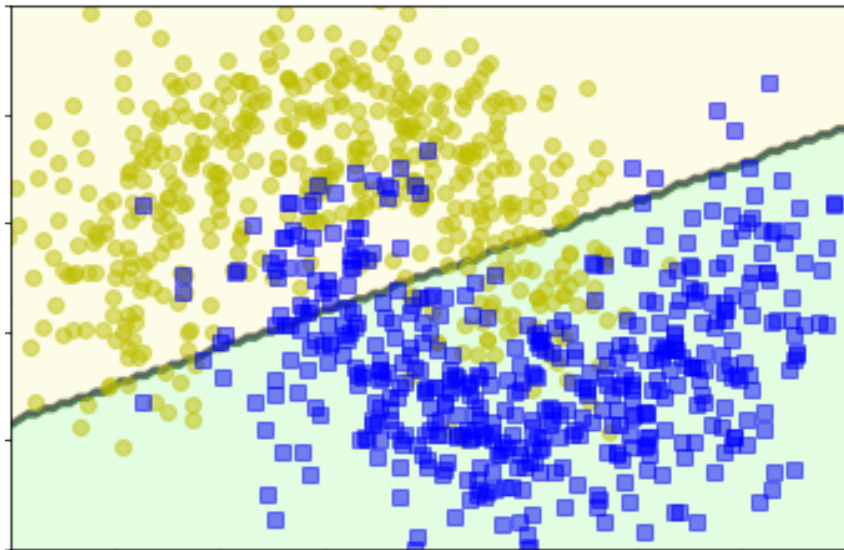
$$\mathbf{x}_i \text{ negative } (y_i = -1): \quad \mathbf{x}_i \cdot \mathbf{w} + b \leq -1$$

$$\text{For support vectors,} \quad \mathbf{x}_i \cdot \mathbf{w} + b = \pm 1$$

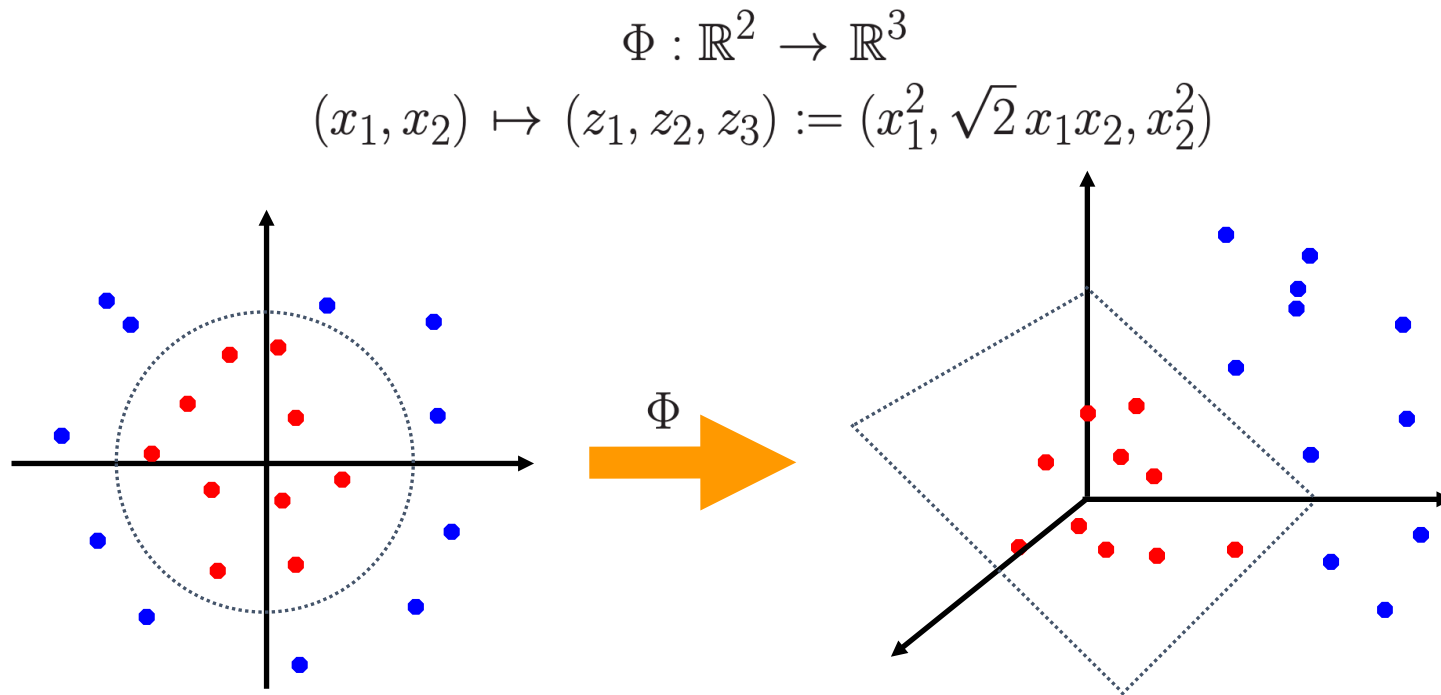
$$\text{Distance between point and hyperplane:} \quad \frac{|\mathbf{x}_i \cdot \mathbf{w} + b|}{\|\mathbf{w}\|}$$

$$\text{Therefore, the margin is } 2 / \|\mathbf{w}\|$$

Real data are non-linear

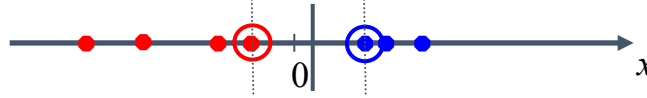


Map the original input space to some higher-dimensional feature space where the training set is separable:



Nonlinear SVMs

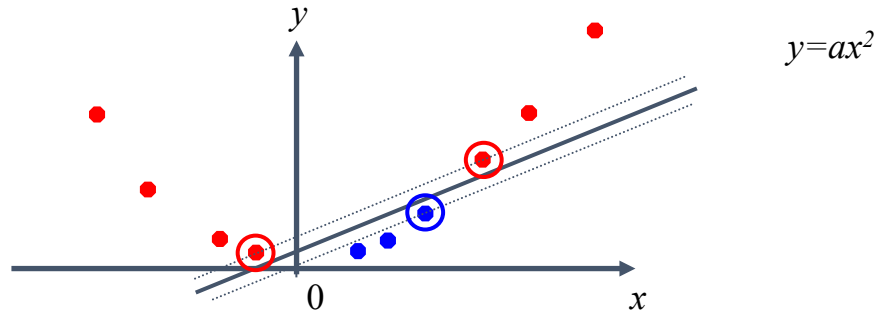
- Linearly separable dataset in 1D:



- Non-separable dataset in 1D:



- We can map the data to a higher-dimensional space:



- **General idea:** the original input space can always be mapped to some higher-dimensional feature space where the training set is separable
- **The kernel trick:** instead of explicitly computing the lifting transformation $\varphi(\mathbf{x})$, define a kernel function K such that

$$K(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}) \cdot \varphi(\mathbf{y})$$

- K is called the Gram matrix, which is positive semidefinite

- Bag of visual words (BoVW) model for image classification
- Different steps of BoVW
- Spatial pyramids histogram for BoVW
- NN classifier and SVM