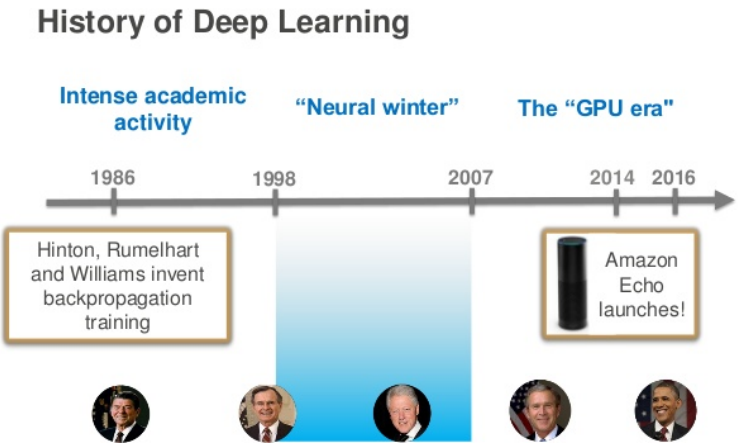
Neural Networks and Deep Learning: Deep Learning Winter History

Nicolas Thome

Conservatoire National des Arts et Métiers (Cnam)
Département Informatique

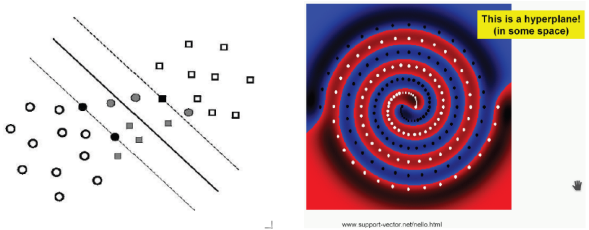
90's: start of 2^{nd} winter for deep learning



Source: Amazon

90's: start of 2^{nd} winter for deep learning

- Deep neural nets =' black magic', black boxes
  - Lack of interpretability
  - Optimization issues for highly non-convex objective function
- **Golden age of kernel methods**
  1. Generalization theory with Support Vector Machines
  2. Extension to non-linear modes: kernel trick
  3. Convex optimization problem



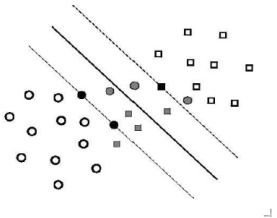
Kernel Methods

1. Generalization Theory with Support Vector Machines (SVM)
   - SVM for binary classification: ~ formal neuron with hinge loss
   - ℓ_2 regularization term $||\mathbf{w}||^2$ ~ weight decay:

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{N} \max\left[0, 1 - \left(\mathbf{w}^\top \mathbf{x}_i + b\right) y_i^*\right]$$

   - Geometric interpretation:
     $1/||\mathbf{w}||^2$ margin between $\oplus/\ominus$ classes
   - SVM Generalization bounds on test error E_{true}:

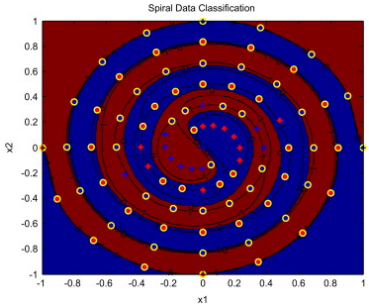
$$E_{true} \leq E_{train} + \phi\left(\frac{h}{N}\right) \quad \text{N \# ex, } h \text{ VC dimension}$$



⇒ **Strong point: theoretical guaranty for SVMs**

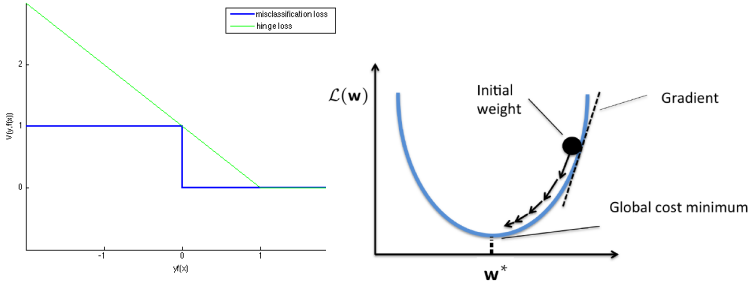
Kernel Methods

2. SVM limited to linear boundaries ⇒ Kernel Trick:
   - Project input data with non-linear injection function ϕ
   - Solve linear SVM in the induced space
- Kernel Trick: no explicitly computation in induced space required
   - Infinite induced spaces possibles, *e.g.* Gaussian kernel



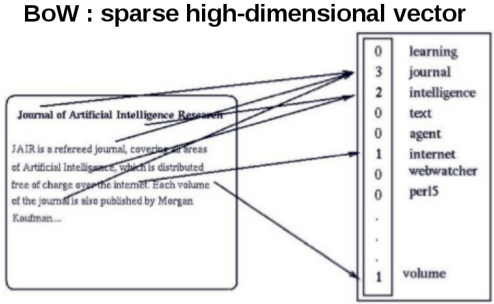
Kernel Methods

3. Even with non-linear kernel ⇒ **convex optimization problem!**
   - Efficiently solved, Global minimum



2000's: Bag of Words Model (BoW)

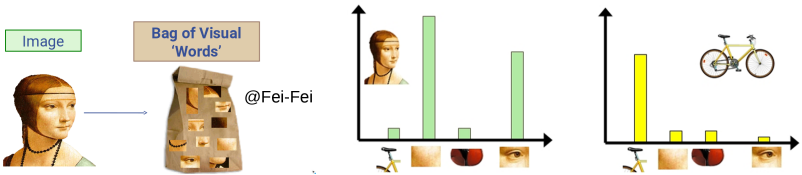
- Started from the Information Retrieval (IR) community
- Text classification: document as a histogram of word occurrences



- Bow representation as input
  for powerful classifiers, *e.g.* SVM

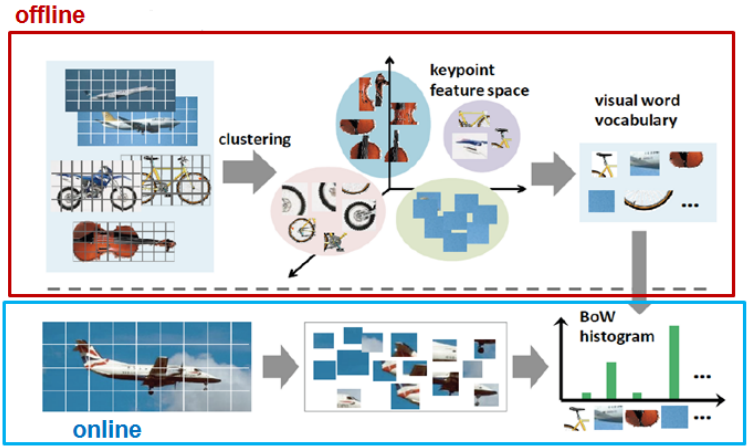
2000's: Bag of Words Model

- Adapting the BoW model for visual recognition ?
  $\Rightarrow$ Bag of Visual Word (BoV)
- Main challenge: definition of visual words unclear!



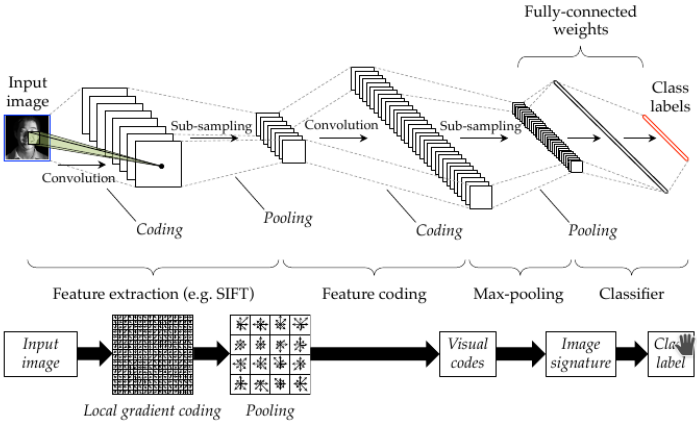
- Solution: compute a dictionary on local image regions (clustering)
  - Local regions represented by handcrafted descriptors, *e.g.* SIFT

2000's: Bag of Visual Words Model



- 2000's: BoW + SVM state-of-the-art
- Many works on kernel on BoW, coding & pooling

BoW *vs* ConvNet



- BoW architecture: ~ 2 [Convolution/Pooling] blocks!
- ConvNet: learned features, deeper hierarchies?
- **BUT:** not enough training data in the 2000's !

Deep Learning Winter: Conclusion

- Decline of Deep Neural Nets due to:
  - Elegant convex competitors (SVM)
    with non-linear boundaries (although not deep)
  - Strong handcrafted feature for important applications,
    e.g. text, image, speech outperforming deep models
- **More data and more power: modern deep learning**
  ⇒ following!

IMAGENET