Tensor Ensemble Learning

Authors:
Ilia Kisil, Ahmad Moniri, Danilo P. Mandic

Electrical and Electronic Engineering Department
Imperial College London, SW7 2AZ, UK

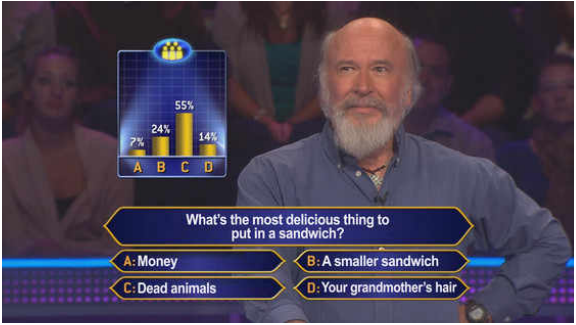
E-mails: {i.kisil15, ahmad.moniri13, d.mandic}@imperial.ac.uk

November 28, 2018

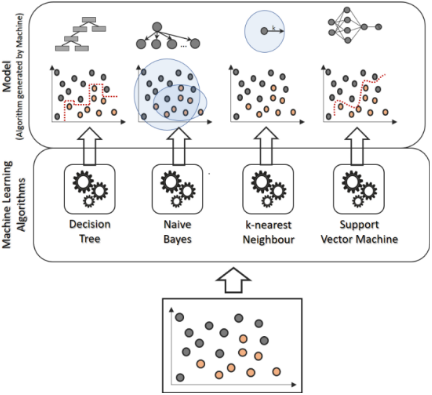
Outline

- Wisdom of the crowd
- Ensemble learning and existing algorithms
- Multidimensional representaiton of data
- Basics of tensor decompositions
- Tensor Ensemble Learning (TEL)
- Simulations and results

Ensemble learning: The wisdom of the crowd



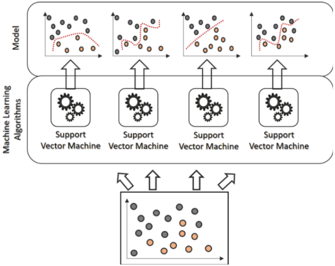
Ensemble learning: Motivation and limitations



- Every model has its own weaknesses ↬ Combining different models can find a better hypothesis
- Every model explores its own hypothesis space ↬ Robust to outliers
- Strong assumption that our individual errors are uncorrelated

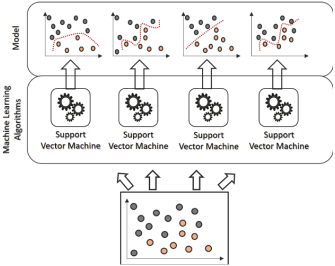
Ensemble learning: Existing approaches

Bagging

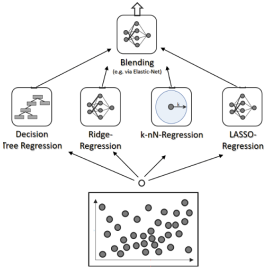


Ensemble learning: Existing approaches

Bagging

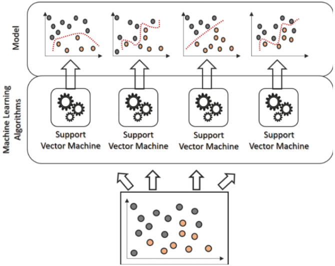


Stacking

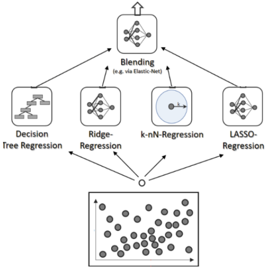


Ensemble learning: Existing approaches

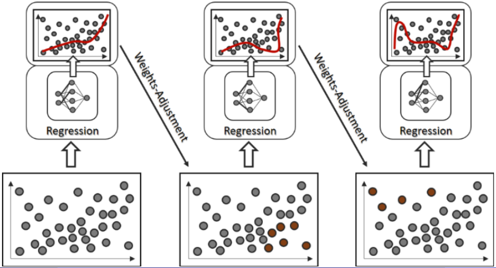
Bagging



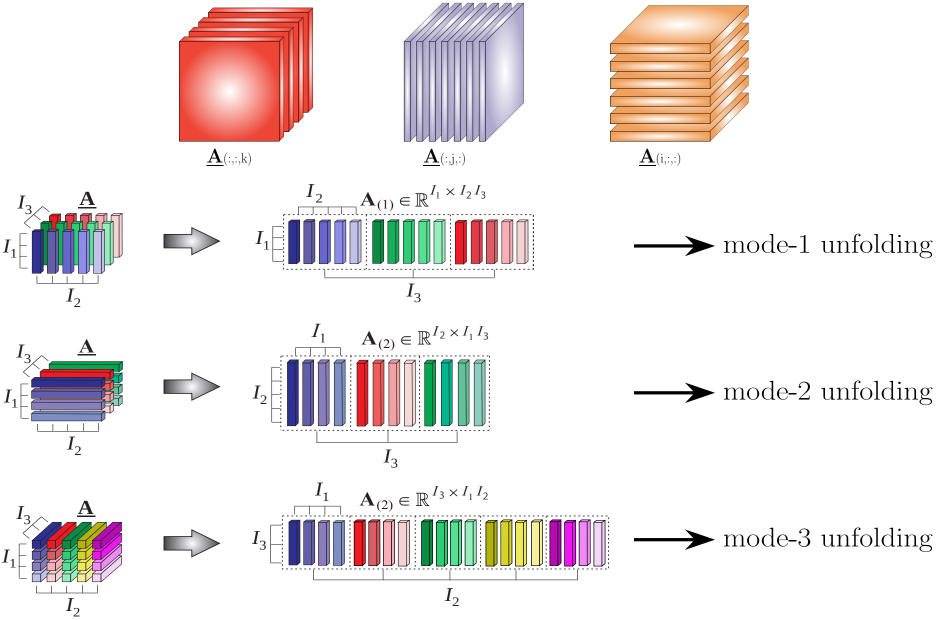
Stacking



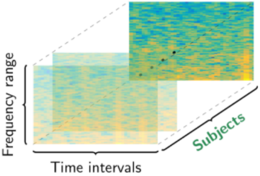
Boosting

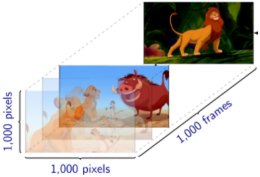


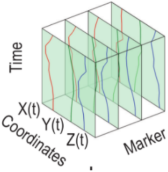
Tensors and basic sub-structures

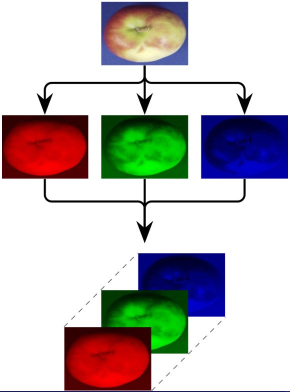


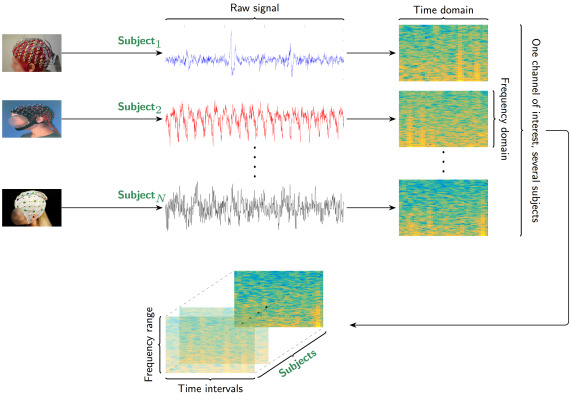
Multidimensional data and tensor construction



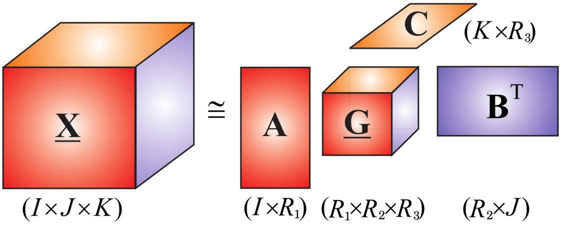






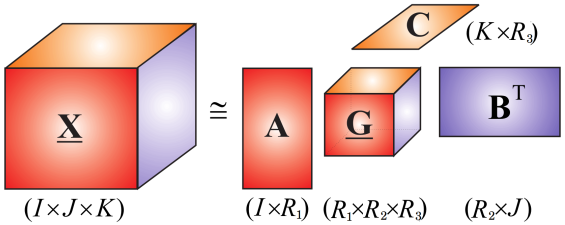


Tucker decomposition ↬ HOSVD



- Eeach vector of $\mathbf{A}$ is associated with every vector of $\mathbf{B}$ and $\mathbf{C}$ through the core tensor $\underline{\mathbf{G}} \looparrowright \underline{\mathbf{X}} \approx \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} \underline{\mathbf{G}}_{r_1 r_2 r_3} \cdot \mathbf{a}_{r_1} \circ \mathbf{b}_{r_2} \circ \mathbf{c}_{r_3}$
- In general, the Tucker decomposition is not unique
- But the subspaces spanned by vectors of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are unique
- By imposing orthogonality constrains on each factor matrix, we arrive at the natural generalisation of the matrix SVD, the higher-order SVD (HOSVD)

Computation of the HOSVD



1. Compute the factor matrices first:

$$\begin{aligned} \mathbf{X}_{(1)} &= \mathbf{A}\mathbf{\Sigma}_{(1)}(\mathbf{V}^{(1)})^T \\ \mathbf{X}_{(2)} &= \mathbf{B}\mathbf{\Sigma}_{(2)}(\mathbf{V}^{(2)})^T \\ \mathbf{X}_{(3)} &= \mathbf{C}\mathbf{\Sigma}_{(3)}(\mathbf{V}^{(3)})^T \end{aligned} \tag{1}$$

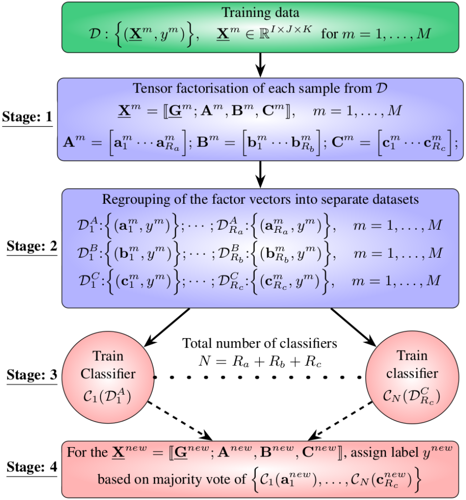
2. Compute the core tensor

$$\underline{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T \tag{2}$$

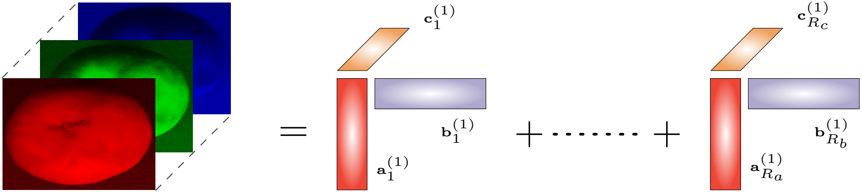
Where $\underline{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{A}^T \Leftrightarrow \mathbf{G}_{(1)} = \mathbf{A}^T\mathbf{X}_{(1)}$

Tensor ensemble learning (TEL): General concept

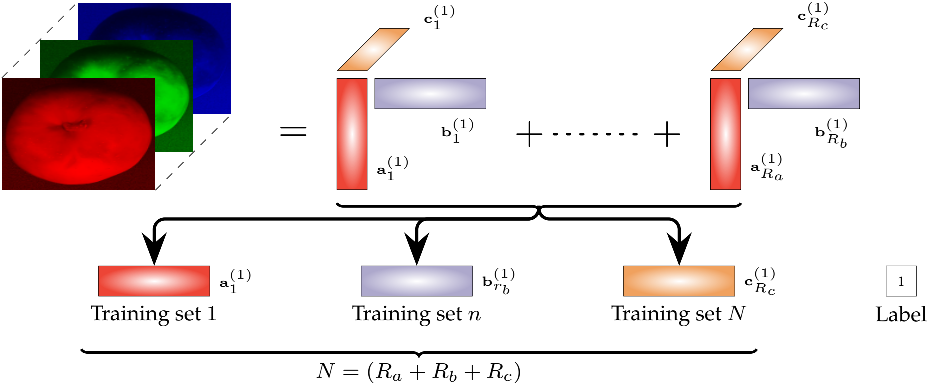
1. Apply tensor decomposition to each multidimensional sample to extract hidden information
2. Perform reorganisation of the obtained latent components
3. Use them to train an ensemble of base learners
4. For new sample, aggregate the knowledge about extracted latent components based on trained models in stage 3



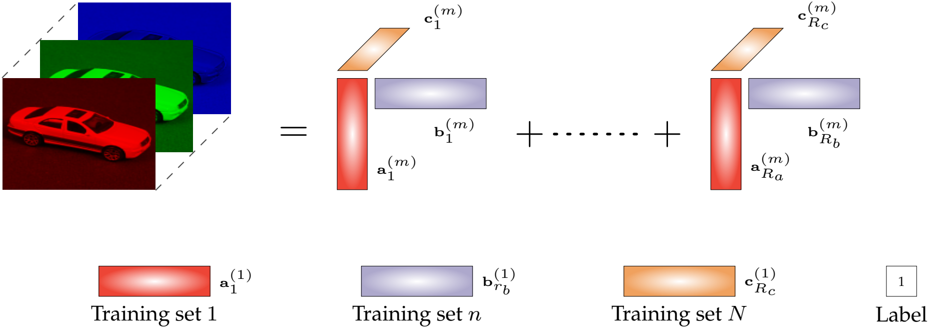
TEL: Formation of training set



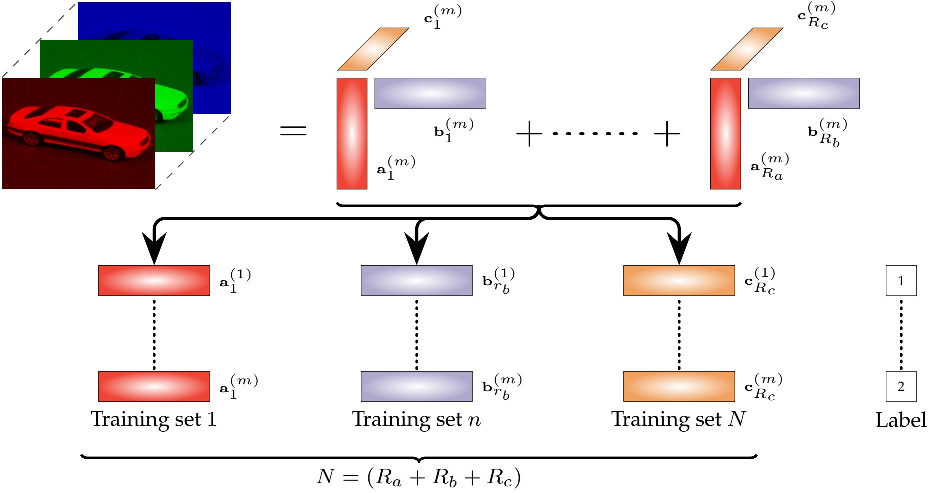
TEL: Formation of training set



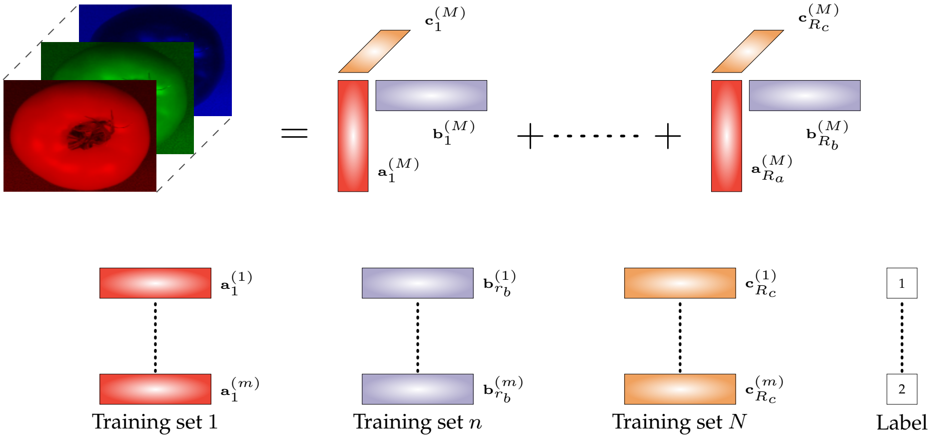
TEL: Formation of training set



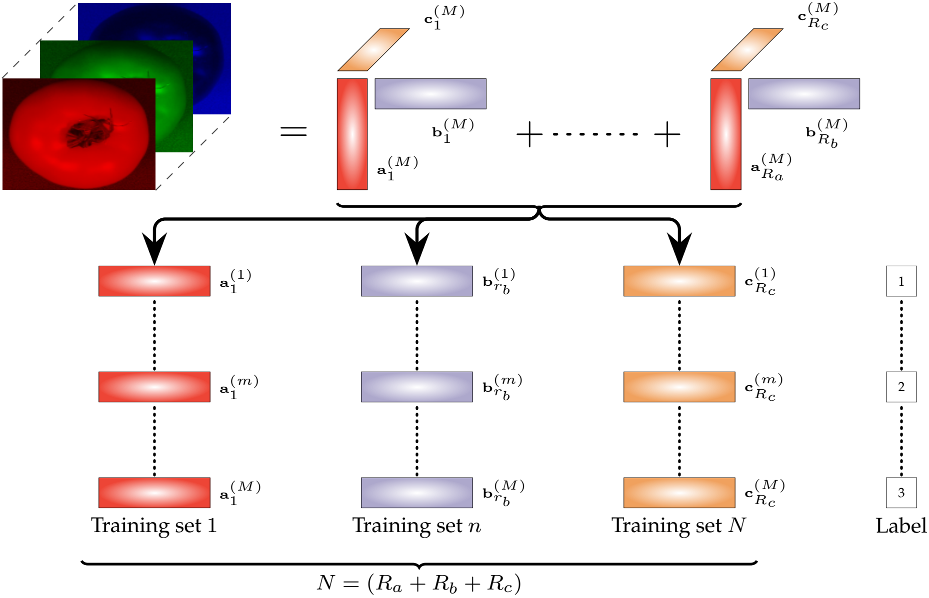
TEL: Formation of training set



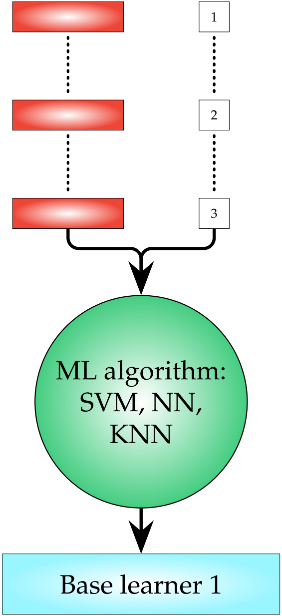
TEL: Formation of training set



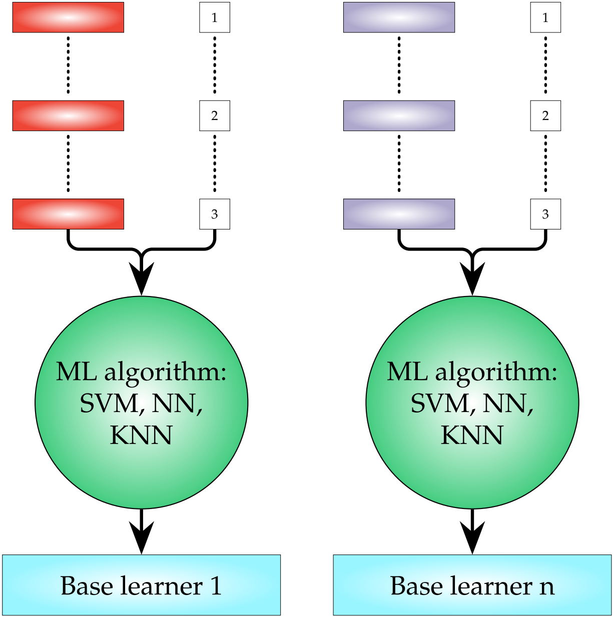
TEL: Formation of training set



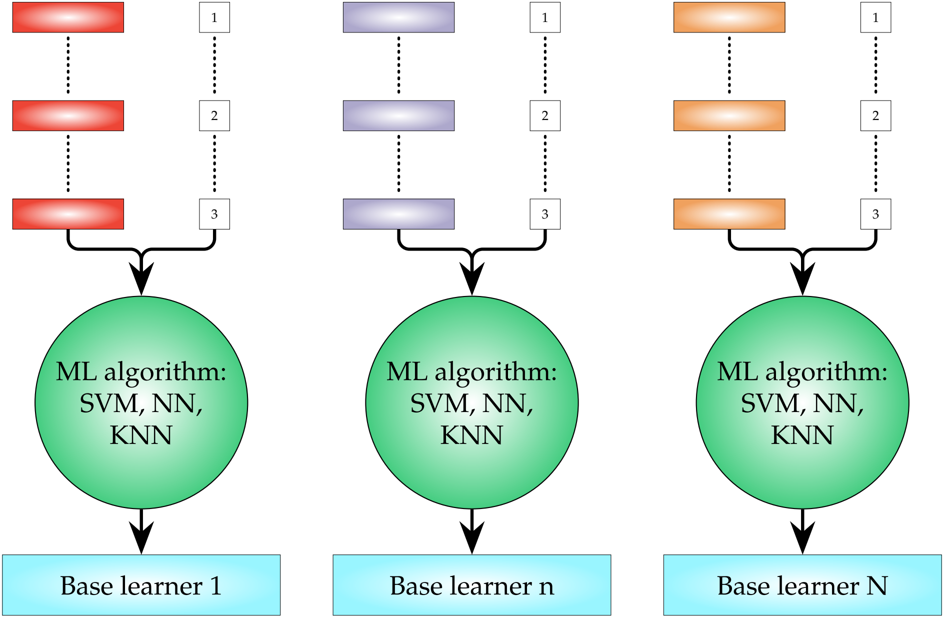
TEL: Training stage



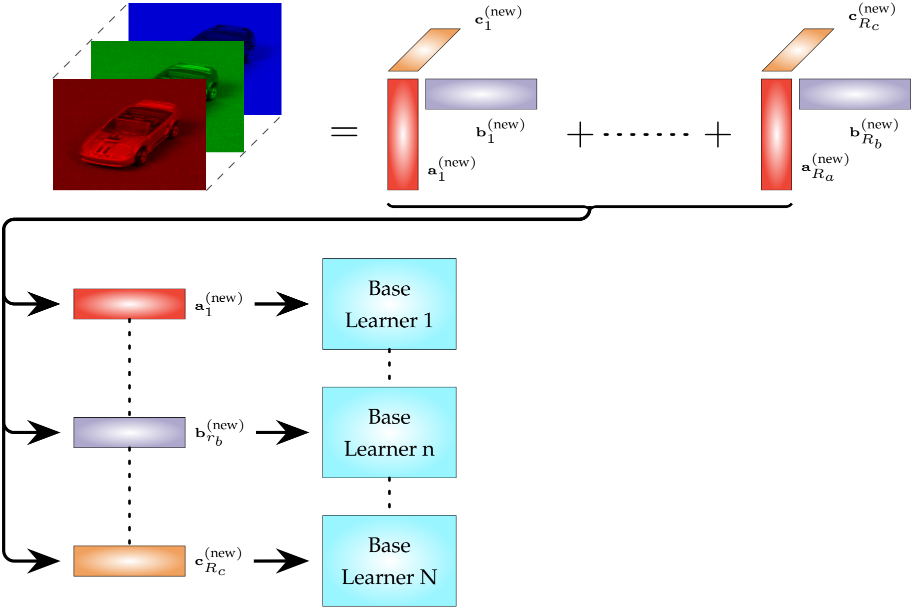
TEL: Training stage



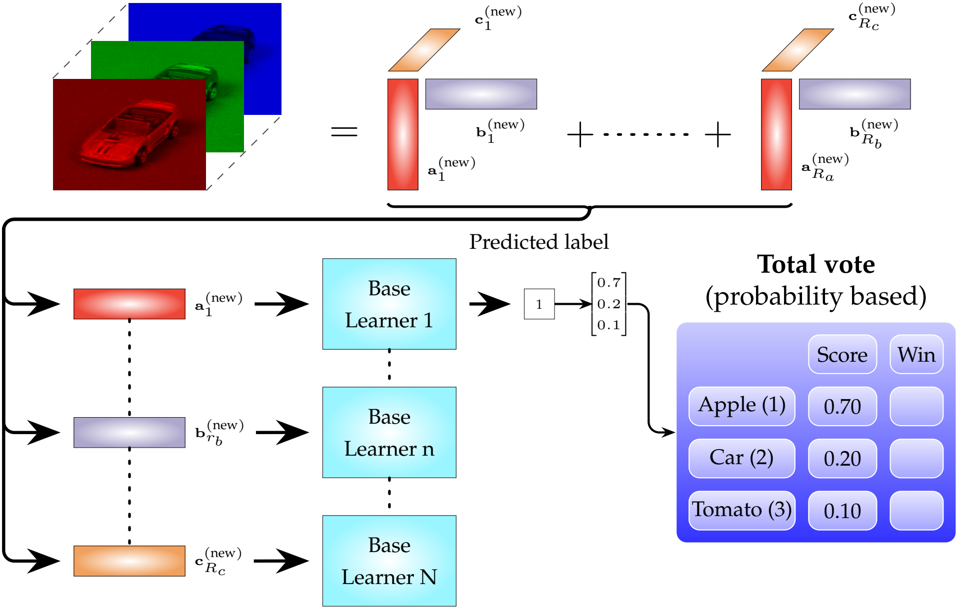
TEL: Training stage



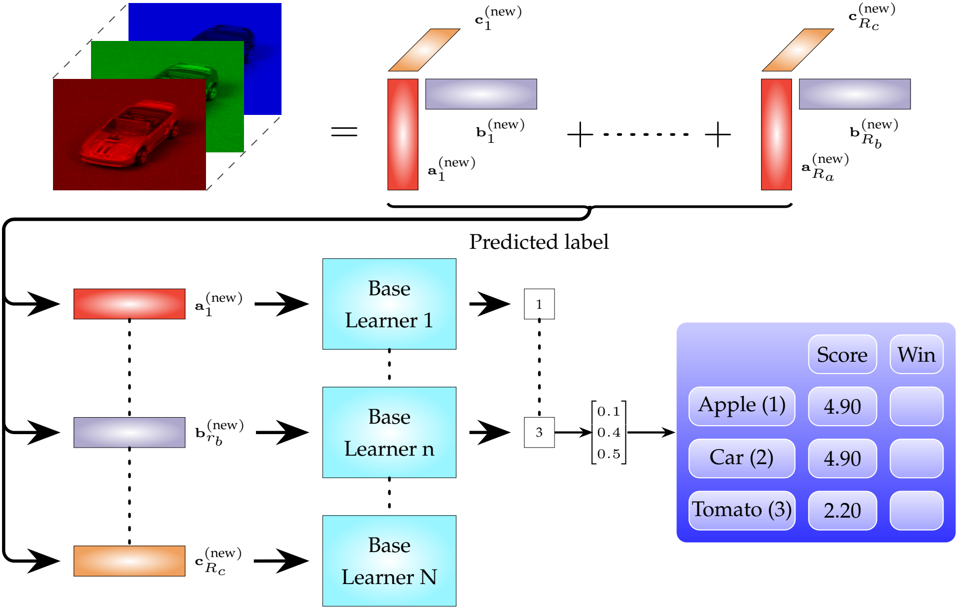
TEL: Testing stage



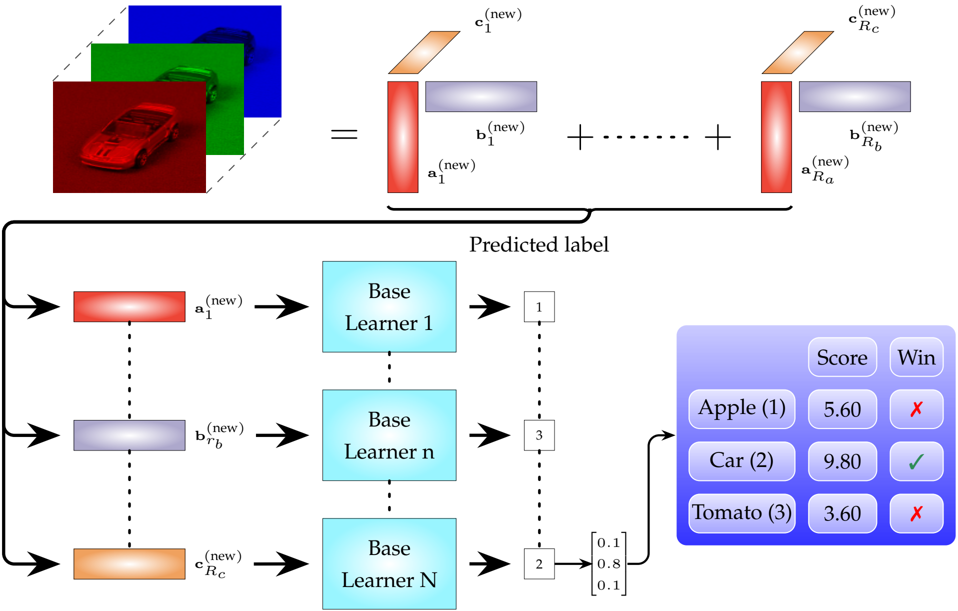
TEL: Testing stage



TEL: Testing stage



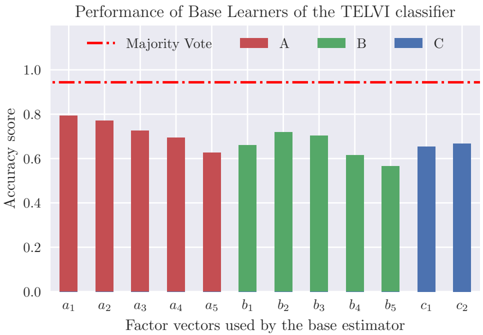
TEL: Testing stage



ETH-80 dataset

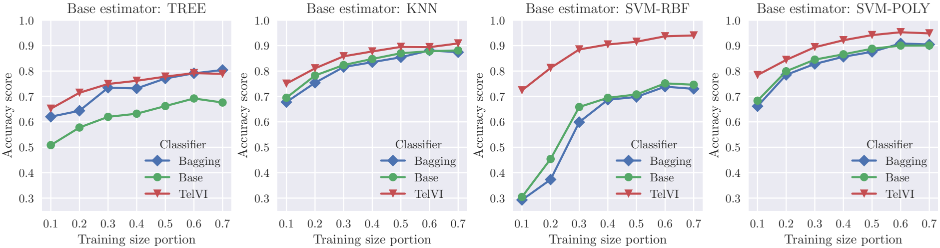
- 8 different categories
- 10 objects per category
- Each object is captured under 41 different viewpoints
- Total: 3280 samples

Performance of base estimators



- We employed HOSVD with multilinear rank $(5, 5, 2)$
- Utilised 12 base estimators (train/test split 50%)
- None of the base classifiers exhibited strong performance on the training set
- Combinational behaviour is similar to classic ensemble learning

Comparison of the overall test performance



- Random split into training and test data varied in range from 10% 70%
- Hyperparameter tuning: grid search with the 5-fold CV of the training data
- For fair comparison, the Bagging classifier also utilised 12 base learners

Conclusions: Key points to take home

1. TEL is a novel framework for generating ensembles of base estimators for multidimensional data
2. TEL highly parallelisable and suitable for large-scale problems
3. Enhanced performance is due to ability to obtain uncorrelated surrogate datasets that are generated by HOSVD

Conclusions: Key points to take home

1. TEL is a novel framework for generating ensembles of base estimators for multidimensional data
2. TEL highly parallelisable and suitable for large-scale problems
3. Enhanced performance is due to ability to obtain uncorrelated surrogate datasets that are generated by HOSVD

New Software: Higher Order Tensors ToolBOX (HOTTBOX)

Our python package for multilinear algebra: `github.com/hottbox/hottbox`

Documentation: `hottbox.github.io`

Tutorials: `github.com/hottbox/hottbox-tutorials`

The End

Thank you for your attention

Questions?