LESSON 24b

Exam simulation

Lessons 14-24




COVID19

- During the COVID19 emergency
  the usual designed exam format is not applicable
- The remote format
  - Will be applied in the 2021 summer session
  - MOODLE+SEB+ZOOM
  - Refer to the specific ARIEL site (Notice Board)
- Nevertheless, with the simulation you have here, you can test your current learning level and accuracy
  - The number of questions is **not** indicative
  - The format of the questions **is** indicative
  - The level of knowledge **is** indicative
  - Try to answer one single question in **1 minute**

Previous simulation

- See the simulation @lesson #4 about questions regarding the first 4 lessons
- See the simulation @lesson #13 about questions regarding lessons 4-13
- In this simulation we will focus on lessons #14-24

Please remember

The remote Multiple-choice test you will find:

- 1 single correct answer → *weigth = 1*
- N wrong answers → *weigth = 0*
- IN THE EXAM no negative values.

1. No penalties for not answering
   →OK BUT IT'S BETTER TO TRY TO ASWER
2. DO NOT LOOK FOR THE RIGHT ANSWER, EXCLUDE THE WRONG ONES
3. DOUBTS? → Choose the answer that "looks less wrong" to you

IMPORTANT! The final score of the exam will be mapped from the score of the test, it is not equal!




EXAM SIMULATION

Lessons #14-24 of the course

To TEST your preparation

(do not use the simulation to study)

On your mark, get set, …




Feature engineering:

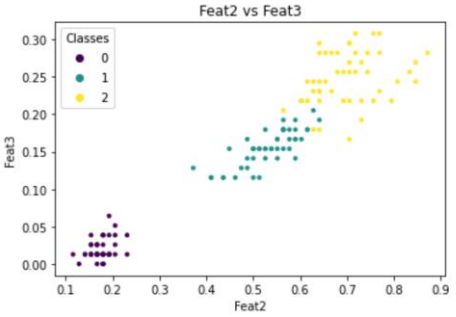
You have a dataset X of 1000 samples and number of features F = 4 features with output labels. You want to reduce the number of features F to 2 for data visualization. According to the goal, consider the following options.

OPTION A: Apply **PCA** to X, and select only the first 2 Principal Components

OPTION B: Apply the **Feedforward Feature Selection** to X, and select only the first 2 more relevant features

1. Option A is possible. Option B is possible.
2. Option A is NOT possible. Option B is possible.
3. Option A is possible. Option B is NOT possible.
4. Option A is NOT possible. Option B is NOT possible.

Feature engineering:



You have a dataset X of 1000 samples and number of features F = 4 features with output labels. You want to reduce the number of features F to 2 for data visualization. According to the goal, consider the following options.

OPTION A: Apply **PCA** to X, and select only the first 2 Principal Components

OPTION B: Apply the **Feedforward Feature Selection** to X, and select only the first 2 more relevant features

1. Option A is possible. Option B is possible.
2. Option A is NOT possible. Option B is possible.
3. Option A is possible. Option B is NOT possible.
4. Option A is NOT possible. Option B is NOT possible.

On your mark, get set, …

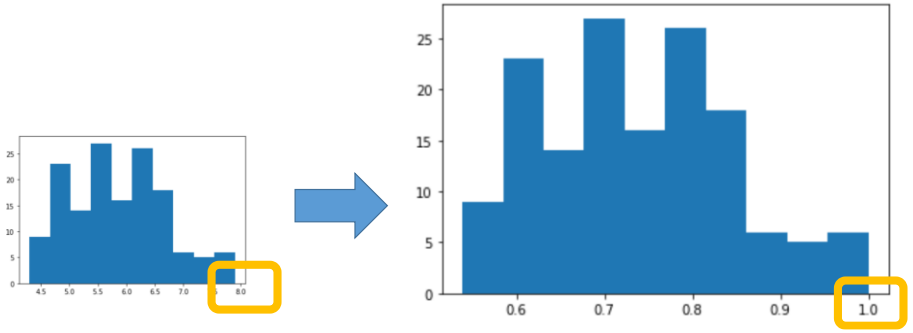



Feature engineering:

You have a feature in your dataset with the following values F1 = [-5 0 +5], which normalization will give you the following F1_norm = [0 0.5 1]

1. Min-MAX
2. Z-score
3. Clipping
4. A different type of normalization

Feature engineering:



You have a feature in your dataset with the following values F1 = [-5 0 +5], which normalization will give you F1_norm = [0 0.5 1].

- *scaling to a range* $x' = (x - x_{min})/(x_{max} - x_{min})$
- *clipping*
- *log scaling* $x' = log(x)$
- *z-score* $x' = (x - \mu)/\sigma$

1. **Min-MAX**
2. Z-score → it's keeping the negative part
3. Clipping → it's not, please see the central value 0→0.5
4. A different type of normalization

On your mark, get set, …

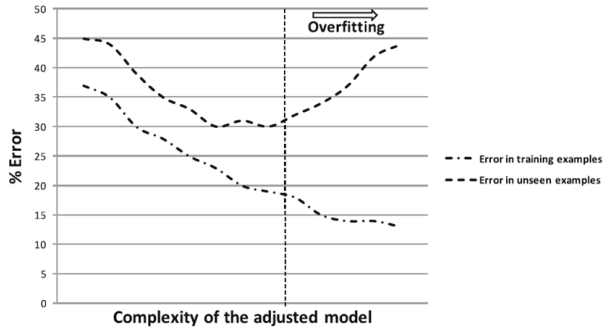



Training:

According to the class discussion, in general for a given small dataset X with output labels, if you train a feed-forward neural models (of the same type) with an increasing number of neurons,
which case is more probable?

1. None of the below
2. The training error and the validation will decrease indefinitely
3. The training error will increase
4. The validation error will decrease indefinitely

Training:



According to the class discussion, in general for a given small dataset X with output labels, if you train a feed-forward neural models (of the same type) with an increasing number of neurons, which case is more probable?

1. None of the below (overfitting can't be excluded!)
2. The training error and the validation will decrease indefinitely
3. The training error will increase
4. The validation error will decrease indefinitely

On your mark, get set, …




Training:

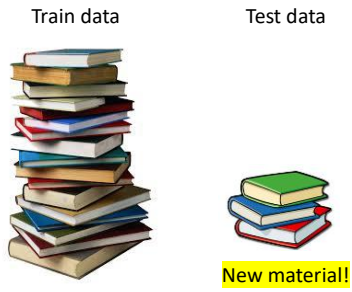
According to the class discussion, in a cross-validation **single** test, which train/test partition of the samples will provide the **lower training** error but the **lower confidence** in the test results?

1. Training set = 99% , Test Set = 01%
2. Training set = 75% , Test Set = 25%
3. Training set = 50% , Test Set = 50%
4. Training set = 25% , Test Set = 75%
5. Training set = 01% , Test Set = 99%

Training:

According to the class discussion, in a cross-validation **single** test, which train/test partition of the samples will provide the **lower training** error but the **lower confidence** in the test results?

1. Training set = 99% , Test Set = 01%
2. Training set = 75% , Test Set = 25%
3. Training set = 50% , Test Set = 50%
4. Training set = 25% , Test Set = 75%
5. Training set = 01% , Test Set = 99%



On your mark, get set, …




Training:

According to the class discussion, what kind of activity can be performed on the **test set**?

1. All the below
2. Mean test error estimation
3. Mean test error estimation and standard deviation
4. Confusion matrix test

Training:

According to the class discussion, what kind of activity can be performed on the **test set**?

1. All the below
2. Mean test error estimation
3. Mean test error estimation and standard deviation
4. Confusion matrix test

On your mark, get set, …




Training:

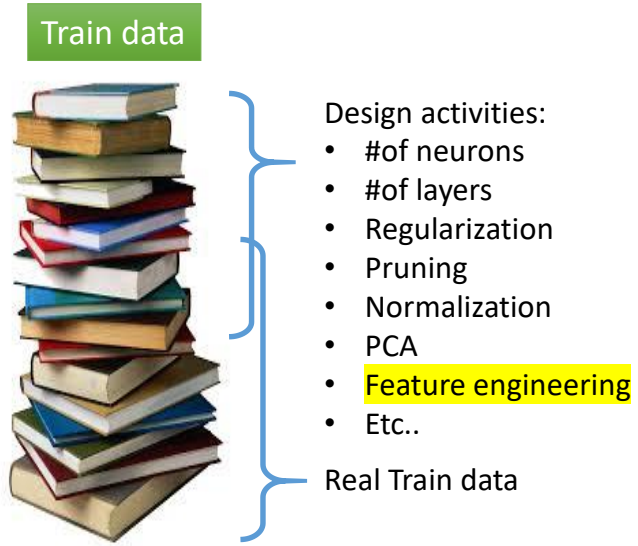
According to the class discussion, what kind of activity can be performed on the **train set**?

- All the other options
- Design of the #of neurons
- Design of the #of layers
- Normalization
- PCA

Training:

According to the class discussion, what kind of activity can be performed on the **train set**?

- All the other options
- Design of the #of neurons
- Design of the #of layers
- Normalization
- PCA



On your mark, get set, …

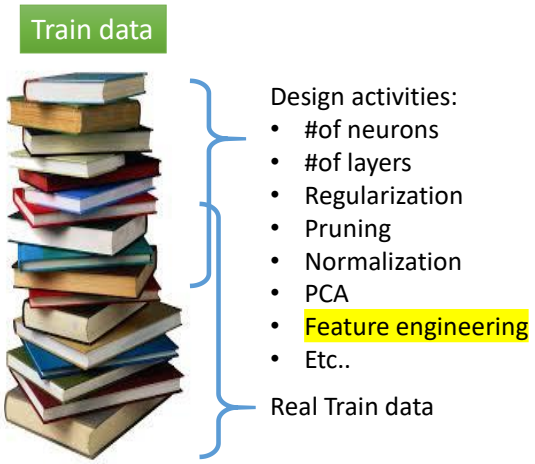



Feature engineering:

According to the class discussion, where can be performed the feature engineering?

- Only on the train set
- Only on the test set
- On the train set and the test set
- Not on the train, Not on the test set, but only on a different dataset

Feature engineering:



According to the class discussion, where can be generally performed the feature engineering?

- Only on the train set
- Only on the test set
- On the train set and the test set (this part is **very** wrong),
- Not on the train, Not on the test set, (this part is wrong), but only (this part is wrong) on a different dataset

On your mark, get set, …




Performance assessment:

A simple *k*-Fold Cross Validation procedure may

- Lead to disarranging the proportion of examples from each class in the test partitions
- Making impossible to process the test error
- Get stuck into one the local minima
- Produce severe overfitting
- None of the other answers

Performance assessment:

A simple *k*-Fold Cross Validation procedure may

- Lead to disarranging the proportion of examples from each class in the test partition
- Making impossible to process the test error
- Get stuck into one the local minima (!)
- Produce severe overfitting (!)
- None of the other answers

Y = [1 1 1 1 1 1 1 1 1 1 2 2 2 2 3 3 3 4 4]

→ 2-Fold Cross Validation

[1 1 1 1 1 1 1 1 1] [1 2 2 2 2 3 3 3 4 4]

On your mark, get set, …

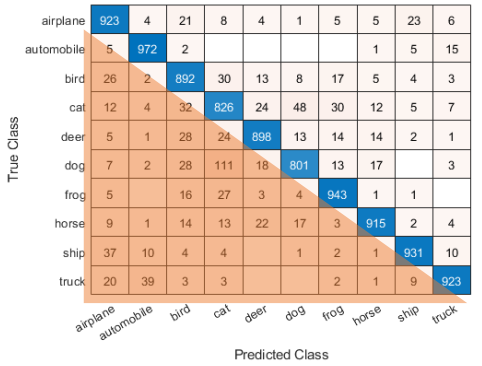



Performance assessment:

Which option is correct?

- From the confusion matrix is possible to process the classification error
- From the confusion matrix is possible to process the classification error and vice versa
- The confusion matrix is applicable only to binary classification systems
- The classification error is equal to the sum of the diagonal elements of the confusion matrix

Performance assessment:



Which option is correct?

- From the confusion matrix is possible to process the classification error
- From the confusion matrix is possible to process the classification error and vice versa
- The confusion matrix is applicable only to binary classification systems
- The classification error is equal to the sum of the diagonal elements of the confusion matrix

On your mark, get set, …




Classification:

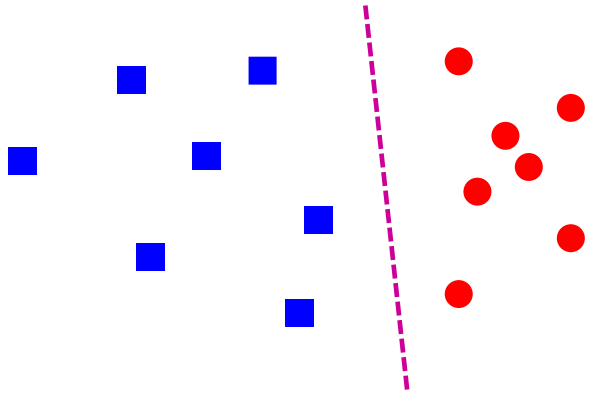
According to the notation used in class, which kind of a model is described by the equation
$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$

- Liner classifier
- Liner regressor
- Soft-max neuron
- Sigmoidal neuron
- Gradient descent formula
- Number of the model's parameters

Classification:

According to the notation used in class, which kind of a model is described by the equation $f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$

- Linear classifier
- Linear regressor
- Soft-max neuron
- Sigmoidal neuron
- Gradient descent formula
- Number of the model's parameters



On your mark, get set, …



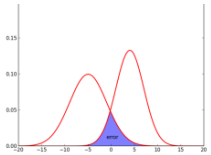

Learning:

According to the notation used in class, which kind of a classifier is better described by the following definition: “the output is the label produced by the most probable classifier”

- Bayes Optimal Classifier
- Supervised Classifier
- K-means
- None of the other options

Learning:

Bayes Optimal Classification



- *Defined as the label produced by the most probable classifier*

$$\arg\max_{v_j \in V} P(v_j|D) = \arg\max_{v_j \in V} \sum_{h_i \in H} P(v_j|h_i)P(h_i|D)$$

(Formula not in the exam)

According to the notation used in class, which kind of a classifier is better described by the following definition: “the output is the label produced by the most probable classifier”

- Bayes Optimal Classifier
- Supervised Classifier
- K-means
- None of the other options

On your mark, get set, …




Learning:

According to the class discussion the kNN classifier, what kind of learning is it?

- Instance-based Learning
- Eager Learning
- Hard-limited Learning
- Unsupervised Clustering
- None of the other options

Learning:

According to the class discussion the kNN classifier, what kind of learning is it?

- Instance-based Learning
- Eager Learning
- Hard-limited Learning
- Unsupervised Clustering
- None of the other options

On your mark, get set, …




Learning:

According to the class discussion, what is the classifier with the following properties:
1) not based on neural techniques;
2) It's deterministic with no random initialization;
3) Perfect repeatability;
4) A minimum number of parameters is needed;
5) Learning is very simple ***but effective****;*
6) Perfect explainability

- kNN
- Linear classifier
- Decision Tree
- K-means
- None of the other options

Learning:

According to the class discussion, what is the classifiers with the following properties:
1) not based on neural techniques;
2) It's deterministic with no random initialization;
3) Perfect repeatability;
4) A minimum number of parameters is needed;
*5) Learning is very simple **but effective**;*
6) Perfect explainability

- kNN
- Linear classifier
- Decision Tree
- K-means
- None of the other options

On your mark, get set, …

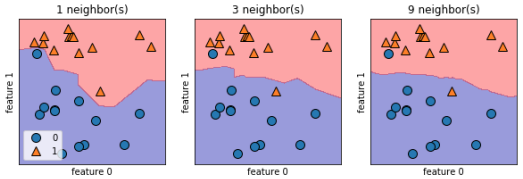


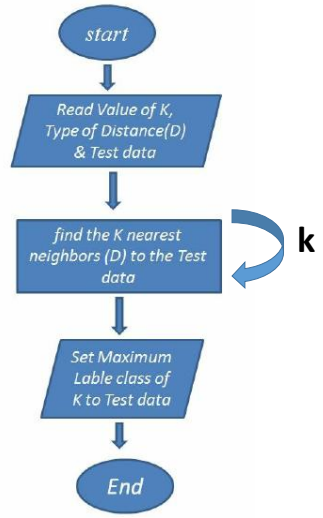

Classifiers:

According to the class discussion on kNN classifiers about the k parameter and its relationship to regularization of the decision boundaries and the computational complexity, what is the correct option about ***larger values*** *of k?*

- More regulatization and more complexity
- Less regulatization and more complexity
- More regulatization and less complexity
- Less regulatization and less complexity
- The parameter k is not related to regularization and complexity

Classifiers:





*According to the class discussion on kNN classifiers about the k parameter and its relationship to regularization of the decision boundaries and the computational complexity, what is the correct option about **larger values** of k?*

- More regulatization and more complexity
- Less regulatization and more complexity
- More regulatization and less complexity
- Less regulatization and less complexity
- The parameter k is not related to regularization and complexity

On your mark, get set, …




Dimensionality reduction:

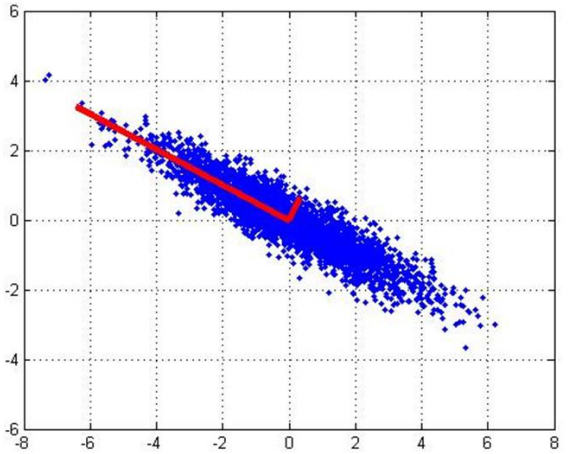
According to the class discussion on PCA what is the correct option?

- PCA vectors are originating from the center of mass of the points
- All subsequent principal component vectors are orthogonal
- All the other options

Dimensionality reduction:

According to the class discussion on PCA what is the correct option?

- PCA vectors are originating from the center of mass of the points
  - All subsequent principal component vectors are orthogonal
- All the other options



On your mark, get set, …




Dimensionality reduction:

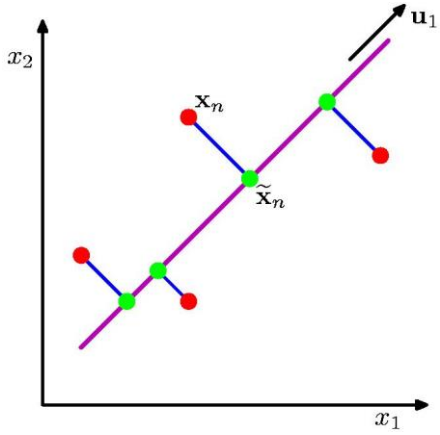
According to the class discussion on PCA what is the correct option?

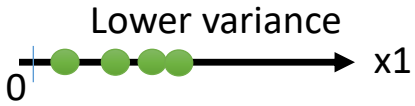
- All subsequent principal component vectors are orthogonal
- The variance of the data projection on the first PCA vectors is maximized
- All the other options

Dimensionality reduction:

According to the class discussion on PCA what is the correct option?

- All subsequent principal component vectors are orthogonal
- The variance of the data projection on the first PCA vectors is maximized
- All the other options





On your mark, get set, …




Unsupervised learning :

According to the class discussion about unsupervised learning , what is the method with the following properties:

- *You need to specify the number of clusters k in advance*
- *Is unable to handle noisy data and outliers*
- *It is not suitable to discover clusters with non-convex shapes*

1. K-means
2. kNN
3. Decision tree
4. None of the other options

Unsupervised learning :

According to the class discussion about unsupervised learning , what is the method with the following properties:

- *You need to specify the number of clusters k in advance*
- *Is unable to handle noisy data and outliers*
- *It is not suitable to discover clusters with non-convex shapes*

1. K-Means
2. kNN
3. Decision tree
4. None of the other options

On your mark, get set, …




Supervised learning :

According to the class discussion, considering the equation of the back-propagation in a feedforward neural network of weight w_ij connected to the following output neuron k, which is the missing term?
*DELTAW_ij = ??? * y_j * delta_k*

1. ??? = alfa (the regularization term < 1)
2. ??? = alfa (the regularization term > 1)
3. ??? = x_j (the input vector)
4. ??? = x_j (the input vector error)

Supervised learning :

According to the class discussion, considering the equation of the back-propagation in a feedforward neural network of weight w_ij connected to the following output neuron k, which is the missing term?
*DELTAW_ij = ??? * y_j * delta_k*

1. ??? = alfa (the regularization term < 1)
2. ??? = alfa (the regularization term > 1)
3. ??? = x_j (the input vector)
4. ??? = x_j (the input vector error)

$$\Delta w_i(p) = \alpha \cdot x_i(p) \cdot e(p)$$

Perceptron

$$\Delta w_{jk}(p) = \alpha \cdot y_j(p) \cdot \delta_k(p)$$

On your mark, get set, …




Convolutional Neural Networks (CNNs)

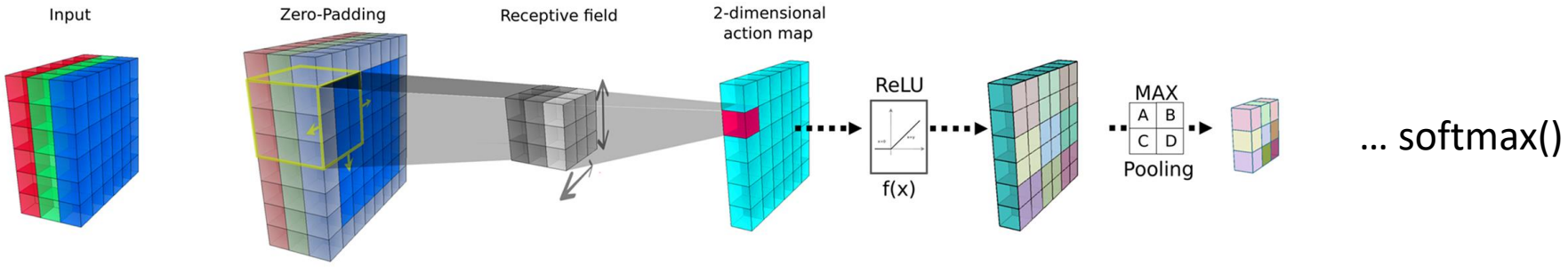
According to the class discussion, considering a general CNN architecture, what is the sequence of modules which is more likely

1. Input layer → Convolution → Relu → Max Pooling … → Softmax → Output layer
2. Input layer → Relu → Convolution → Max Pooling … → Softmax → Output layer
3. Input layer → Relu → Max Pooling Convolution … → Softmax → Output layer
4. Input layer → Relu → Max Pooling→ Softmax→ Convolution … → Output layer

Convolutional Neural Networks (CNNs)

According to the class discussion, considering a general CNN architecture, what is the sequence of modules which is more likely

1. Input layer → Convolution → Relu → Max Pooling … → Softmax → Output layer
2. Input layer → Relu → Convolution → Max Pooling … → Softmax → Output layer
3. Input layer → Relu → Max Pooling Convolution … → Softmax → Output layer
4. Input layer → Relu → Max Pooling→ Softmax→ Convolution … → Output layer



On your mark, get set, …




Industrial Vision Systems

According to the class discussion, considering a standard intelligent vision system, which capability can be processed onboard on a recent smart industrial camera?

1. Segmentation
2. Segmentation, Measurement
3. Segmentation, Measurement, Classification with trained non-deep models
4. Segmentation, Measurement, Classification with trained deep models
5. Segmentation, Measurement, Classification with trained deep models and training of deep models

Industrial Vision Systems

According to the class discussion, considering a standard intelligent vision system, which capability can be processed onboard on a recent smart industrial camera?

1. Segmentation
2. Segmentation, Measurement
3. Segmentation, Measurement, Classification with trained non-deep models
4. Segmentation, Measurement, Classification with trained deep models
5. Segmentation, Measurement, Classification with trained deep models and training of deep models

2018-2020
- Smart cameras with FPGA
- With onboard deep neural network
- Measurements
- OCR + QR + Barcode adv. techniques

As explained, training is typically done off-line and **not** onboard

On your mark, get set, …



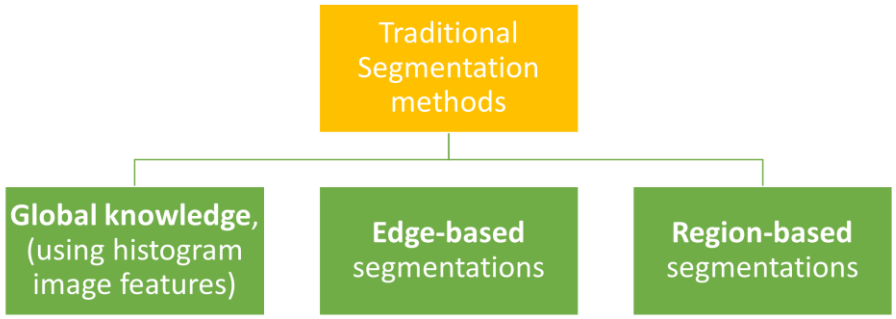

Segmentation

According to the class discussion, Traditional Segmentation methods are quite useful to produce blobs or object candidates to be further processed by deep models for classification or measurements.

Traditional Segmentation methods can be partitioned in

1. Global knowledge, Edge-based
2. Edge-based, Region-based
3. Global knowledge, Edge-based, Region-based
4. None of the other options

Segmentation



According to the class discussion, Traditional Segmentation methods are quite useful to produce blobs or object candidates to be further processed by deep models for classification or measurements.

Traditional Segmentation methods can be partitioned in

1. Global knowledge, Edge-based
2. Edge-based, Region-based
3. Global knowledge, Edge-based, Region-based
4. None of the other options

On your mark, get set, …




HW for deep learning

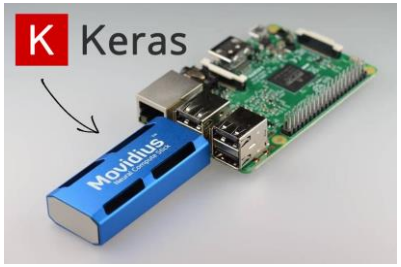
According to the class discussion referred to edge computing, is it possible to process images with trained deep learning models on external small dedicated devices connect via USB connection?

- *True: the usage of dedicated processors and the USB bandwidth make this option possible*
- *False: the USB bandwidth make this option not possible*
- *False: the needed computational complexity needed to run trained deep learning models make this option not possible*
- *False: the bandwidth and the computational complexity need to process images with trained deep learning model is not adequate*

HW for deep learning

According to the class discussion referred to edge computing, is it possible to process images with trained deep learning models on external small dedicated devices connect via USB connection?

- *True: the usage of dedicated processors and the USB bandwidth make this option possible*
- *False: the USB bandwidth make this option not possible*
- *False: the needed computational complexity needed to run trained deep learning models make this option not possible*
- *False: the bandwidth and the computational complexity need to process images with trained deep learning model is not adequate*



E.g., the Intel Movidius Neural Compute Stick (NCS) is a tiny fanless deep-learning device that can be used to learn AI programming at the edge

On your mark, get set, …




Deep learning training

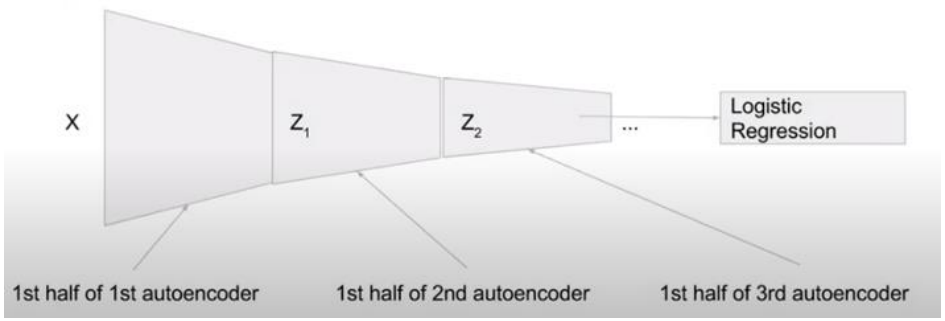
According to the class discussion what is Greedy Layer-Wise Training?

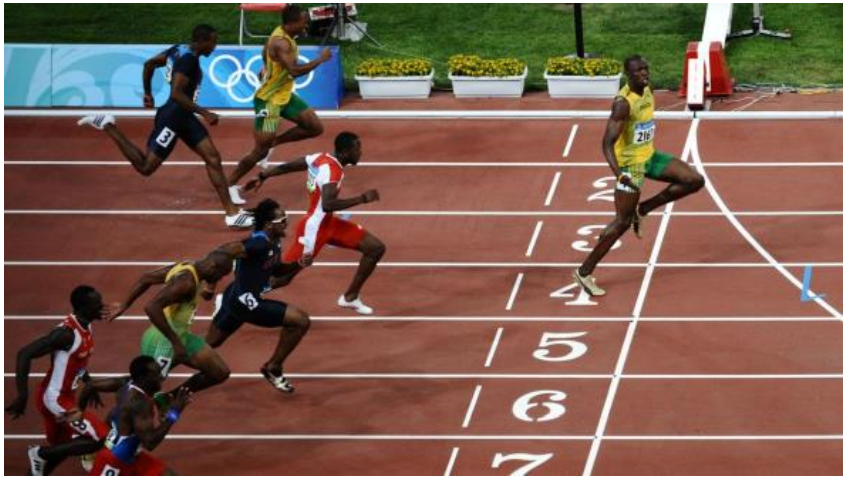
1. *A supervised training step to improve auto-encoders*
2. *A supervised training step to classical feedforward networks*
3. *An unsupervised training step to classical feedforward networks*
4. *An unsupervised training step to improve auto-encoders*

Deep learning training

According to the class discussion, what is Greedy Layer-Wise Training?

1. *A supervised training step to improve auto-encoders*
2. *A supervised training step to classical feedforward networks*
3. *An unsupervised training step to classical feedforward networks*
4. *An unsupervised training step to improve auto-encoders*





How is your accuracy?

Do not hesitate to contact the teacher in case of doubts




What about topics not present in the simulation?

- They are equally important
  - The absence of topics in the simulation does not mean they are not relevant
- You can find them in the exam
- Use the simulation not to study but to understand you are profitably attending the course and, in case, change the study method
- Try to create your own simulation by browsing all slides of the course and ask yourself: what can be asked in this slide?

- **IMPORTANT! The final score of the exam will be mapped from the score of the test, it is not equal!**

Thank you for your attention!

Do not hesitate to contact me in case of doubts

fabio.scotti@unimi.it