
Lab 3: EM Algorithm

Release 1.0

Zhankun Luo

Nov 08, 2021

CONTENTS

1	API Documentation	1
----------	--------------------------	----------

API DOCUMENTATION

`data.mk_data.choose_data` ($X, \pi, se=0$)
Choose data samples from different Gaussian distributions with prior weights

`data.mk_data.generate_data` ($R, \mu, N, se=0$)
Generate N data samples for a Gaussian distribution cluster

`data.mk_data.plot_data` ($data$)
Plot the data generated by the Gaussian Mixture Model

`data.mk_data.save_data` ($data, label$)
Save both data and labels of the Gaussian Mixture Model

`src.utils.circle_cluster` ($R, \mu, category, linestyle='solid'$)
Draw the circle of Gaussian distribution

`src.utils.plot_cluster_center` ($k, mean$)
Plot the center of a cluster

`src.utils.plot_cluster_center_merged` ($l, m, k, mean_l, mean_m, mean_merged$)
Plot the center of the merged cluster and centers of old two clusters

`src.utils.plot_data_label` ($data, label$)
Plot both data and labels

`src.utils.plot_iteration_md1` ($iteration_md1$)
Plot the MDL values for each EM iterations and after merging clusters

`src.utils.plot_list_md1` ($list_md1$)
Plot the minimal MDL values for each number of clusters K

class `src.GMM.GaussianMixture` ($n_clusters=1, tol=None, max_iter=1000, weights_init=None, means_init=None, precisions_init=None, random_state=None, ignore_converged=False$)

__init__ ($n_clusters=1, tol=None, max_iter=1000, weights_init=None, means_init=None, precisions_init=None, random_state=None, ignore_converged=False$)
 Initialize self. See `help(type(self))` for accurate signature.

_compute_log_det_cholesky ($matrix_chol, n_features$)
 Compute the logarithm determinant of cholesky matrix of inverse covariance

_compute_precision_cholesky ($covariances$)
 Compute the cholesky matrix of the inverse covariance

_e_step (X)
 E step of EM algorithm

`_estimate_gaussian_params` (*X*, *resp*)
Estimate new parameters based on the responsibilities $p(Z | X)$

`_estimate_log_gaussian_prob` (*X*, *means*, *precisions_chol*)
Compute the PDF $f(X | Z)$

`_estimate_log_prob` (*X*)
Estimate the log-probabilities, $\log f(X | Z)$

`_estimate_log_prob_resp` (*X*)
Return the logarithm of $f(X)$ and $p(Z | X)$

`_get_iteration_md1` ()
Get the MDL values for all the iterations

`_get_parameters` ()
Get the parameters of Gaussian Mixture Model

`_initialize_params` (*X*)
Initialize parameters including weights, means, covariances

`_m_step` (*X*, *log_resp*)
M step of EM algorithm

`_n_parameters` ()
Return the number of free parameters in the model

`_set_parameters` (*params*)
Set the parameters of Gaussian Mixture Model

`aic` (*X*)
Return AIC criteria

`fit` (*X*)
Fit the model with data using EM algorithm

`mdl` (*X*)
Return MDL criteria

`predict` (*X*)
Predict labels based on the trained parameters

`predict_prob` (*X*)
Predict the probability $p(Z | X)$ based on trained parameters

`score` (*X*)
Compute $\log f(X | \sigma)$ log-likelihood of the given data *X*.

`class src.GMM.OrderIdentification` (*weights*, *means*, *covariances*, *precisions_cholesky*)

`__init__` (*weights*, *means*, *covariances*, *precisions_cholesky*)
Initialize self. See `help(type(self))` for accurate signature.

`_compute_distance_cluster` (*l*, *m*)
Compute the average distance $d(l, m) / N$ of the *l*-th, *m*-th clusters

`_compute_merged_params` (*l*, *m*)
Computet the parameters of the merged cluster based *l*-th, *m*-th clusters

`_get_index_cluster_merged` ()
Get the indices of clusters used to form a new merged cluster

_get_parameters ()

Get the parameters of clusters

_get_params_merged ()

Get the parameters of the merged cluster

merge ()

Merge the l-th, m-th clusters