

(23A31401T) MACHINE LEARNING

Assignment-I

Q1. List out the names and formulas of proximity measures, distance measures, similarity measures, nearest neighbor classification and regression algorithms, and performance evaluation metrics used in machine learning.

	Category	Name	Formula
Ex	Distance Measure	Euclidean Distance	-
	Minkowski Distance		
	Similarity (Non-Metric)	Jaccard Similarity	
	Binary Pattern Proximity		

Get From Text Book Note Down all 1 unit to 3 ½ Unit

Q2. Study the distance and similarity measures available in sklearn.metrics and sklearn.neighbors modules. List out the names of the distance and similarity metrics and write the corresponding mathematical formulas. Identify whether they are used in regression or classification.

So you must:

1. **Filter only distance & similarity metrics**
2. **Ignore displays, plotting, helpers**
3. **Write textbook formulas**
4. **Mention usage (Regression / Classification)**

```
from sklearn import metrics
```

```
import sklearn
```

```
dir(sklearn.metrics.pairwise.distance_metrics)
```

```
['ConfusionMatrixDisplay',  
 'DetCurveDisplay',  
 'DistanceMetric',  
 'PrecisionRecallDisplay',  
 'PredictionErrorDisplay',
```

```
'RocCurveDisplay',
'__all__',
'__builtins__',
'__cached__',
'__doc__',
'__file__',
'__loader__',
'__name__',
'__package__',
'__path__',
'__spec__',
'_base',
'_classification',
'_dist_metrics',
'_pairwise_distances_reduction',
'_pairwise_fast',
'_plot',
'_ranking',
'_regression',
'_scorer',
'accuracy_score',
'adjusted_mutual_info_score',
'adjusted_rand_score',
'auc',
'average_precision_score',
'balanced_accuracy_score',
'brier_score_loss',
'calinski_harabasz_score',
'check_scoring',
'class_likelihood_ratios',
'classification_report',
'cluster',
'cohen_kappa_score',
'completeness_score',
'confusion_matrix',
'consensus_score',
'coverage_error',
'd2_absolute_error_score',
'd2_log_loss_score',
'd2_pinball_score',
'd2_tweedie_score',
'davies_bouldin_score',
'dcg_score',
'det_curve',
'euclidean_distances',
'explained_variance_score',
'f1_score',
'fbeta_score',
'fowlkes_mallows_score',
'get_scorer',
'get_scorer_names',
'hamming_loss',
'hinge_loss',
'homogeneity_completeness_v_measure',
'homogeneity_score',
'jaccard_score',
'label_ranking_average_precision_score',
'label_ranking_loss',
'log_loss',
```

```

'make_scorer',
'matthews_corrcoef',
'max_error',
'mean_absolute_error',
'mean_absolute_percentage_error',
'mean_gamma_deviance',
'mean_pinball_loss',
'mean_poisson_deviance',
'mean_squared_error',
'mean_squared_log_error',
'mean_tweedie_deviance',
'median_absolute_error',
'multilabel_confusion_matrix',
'mutual_info_score',
'nan_euclidean_distances',
'ndcg_score',
'normalized_mutual_info_score',
'pair_confusion_matrix',
'pairwise',
'pairwise_distances',
'pairwise_distances_argmin',
'pairwise_distances_argmin_min',
'pairwise_distances_chunked',
'pairwise_kernels',
'precision_recall_curve',
'precision_recall_fscore_support',
'precision_score',
'r2_score',
'rand_score',
'recall_score',
'roc_auc_score',
'roc_curve',
'root_mean_squared_error',
'root_mean_squared_log_error',
'silhouette_samples',
'silhouette_score',
'top_k_accuracy_score',
'v_measure_score',
'zero_one_loss']

```

Metrics name	ml model Under Name(Like knn or tree,..soon)	formula

Q3. stages involved in a Machine Learning process.?

Q4. Explain the different paradigms of Machine Learning ?

Q5. Critically analyze the limitations of traditional batch machine learning models implemented in **scikit-learn**. How do online learning models in the **River** library overcome these limitations? Explain why River is more suitable for real-time and streaming data applications.(You searching Online Write Short 10 difference Points)