

[Feng2017] W.Feng and M.Chen. Learning Apache Spark, Github 2017.
[Feng2016PSD] W.Feng, A.J.Salgado, C.Wang, S.M.Wise. Preconditioned Steepest Descent Methods for some Nonlinear Elliptic Equations Involving p-Laplacian Terms. J.Comput.Phys., 334:45–67, 2016.
[Feng2014] W.Feng. Prelim Notes for Numerical Analysis, The University of Tennessee, Knoxville.
[Karau2015] H.Karau, A.Konwinski, P.Wendell and M.Zaharia. Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly Media, Inc., 2015
[Kirillov2016] Anton Kirillov. Apache Spark: core concepts, architecture and internals. <http://datas trophic.io/core-concepts-architecture-and-internals-of-apache-spark/>

A

- ~~BinaryLogisticRegressionTraining~~
- ~~Summaryaccuracy(pyspark.ml.classification.LogisticRegressionSummary(class in pyspark.ml.classification))~~
- ~~BisectingKMeans(class in pyspark.ml.clustering)~~
- ~~BisectingKMeansModel(class in pyspark.ml.clustering)~~
- ~~AFTSurvivalRegression(class in pyspark.ml.regression)~~
- ~~BisectingKMeansSummary(class in pyspark.ml.clustering)~~
- ~~AFTSurvivalRegressionModel(class in pyspark.ml.regression)~~
- ~~aic(pyspark.ml.regression.GeneralizedLinearRegressionSummarypark.ml.regression.IsotonicRegressionModelAttribute)~~
- ~~ALS(class in pyspark.ml.recommendation)~~
- ~~build()(pyspark.ml.tuning.ParamGridBuilderALSModel(class in pyspark.ml.recommendation))~~

- BinaryLogisticRegressionTraining
- Summaryaccuracy (pyspark.ml.classification.LogisticRegressionSummary(class in pyspark.ml.classification))
- BisectingKMeans (class in pyspark.ml.clustering)
- BisectingKMeansModel (class in pyspark.ml.clustering)
- AFTSurvivalRegression (class in pyspark.ml.regression)
- BisectingKMeansSummary (class in pyspark.ml.clustering)
- AFTSurvivalRegressionModel (class in pyspark.ml.regression)
- aic
(pyspark.ml.regression.GeneralizedLinearRegressionSummarypark.ml.regression.IsotonicRegressionModelAttribute)
- ALS (class in pyspark.ml.recommendation)
- build() (pyspark.ml.tuning.ParamGridBuilderALSModel (class in pyspark.ml.recommendation))

C

areaUnderROC(pyspark.ml.classification.BinaryLogisticRegressionSummaryChiSquareTest(class in pyspark.ml.stat))

assignClusters() (pyspark.ml.clustering.PowerIterationClusteringmethod), 361

avgMetrics (pyspark.ml.tuning.CrossValidatorModelAttribute), 373

BbaseOn()(pyspark.ml.tuning.ParamGridBuildermethod), 371

bestModel (pyspark.ml.tuning.CrossValidatorModelAttribute), 373

bestModel (pyspark.ml.tuning.TrainValidationSplitModelAttribute), 375

BinaryClassificationEvaluator(class in pyspark.ml.evaluation), 376

BinaryLogisticRegressionSummary(class in pyspark.ml.classification), 332

clusterCenters() (pyspark.ml.clustering.BisectingKMeansModelmethod), 349

clusterCenters() (pyspark.ml.clustering.KMeansModelmethod), 351

ClusteringEvaluator(class in pyspark.ml.evaluation), 378

coefficientMatrix
(pyspark.ml.classification.LogisticRegressionModelAttribute), 329coefficients

(pyspark.ml.classification.LinearSVCModelAttribute), 325coefficients (pyspark.ml.classification.LogisticRegressionModelAttribute), 329coefficients

(pyspark.ml.regression.AFTSurvivalRegressionModelAttribute), 305coefficients

pyspark.ml.classification,324pyspark.ml.clustering,347

pyspark.ml.evaluation,375pyspark.ml.pipeline,368pyspark.ml.recommendation,363

pyspark.ml.regression,303pyspark.ml.stat,297
pyspark.ml.tuning,370

attribute),332
assignClusters() (pys-
park.ml.clustering.PowerIterationClusteringmethod),361
avgMetrics(pyspark.ml.tuning.CrossValidatorModelattribute),373

D
DecisionTreeClassificationModel(classinpyspark.ml.classification),335method),375evaluate() (pyspark.ml.regression.GeneralizedLinearRegressionModelmethod),312evaluate()
(pyspark.ml.regression.LinearRegressionModelmethod),318evaluateEachIteration() (pyspark.ml.classification.GBTClassificationModelmethod),338evaluateEachIteration()
(pyspark.ml.regression.GBTRegressionModelmethod),308Evaluator(classinpyspark.ml.evaluation),375 explainedVariance
(pyspark.ml.regression.LinearRegressionSummaryGBTClassificationModel(class in pysattribute),319 park.ml.classification),338 GBTClassifier
(class in pys- F park.ml.classification),336falsePositiveRateByLabel(pys-
GBTRegressionModel(classinpyspark.ml.classification.LogisticRegressionSummarypark.ml.regression),308attribute),330
GBTRegressor(classinpyspark.ml.regression),featureImportances (pyspark.ml.classification.DecisionTreeClassificationModelGeneralizedLinearRegression(class inattribute),335
pyspark.ml.regression),309 featureImportances (pys- GeneralizedLinearRegressionModel
park.ml.classification.GBTClassificationModel(classinpyspark.ml.regression),312 attribute),338
GeneralizedLinearRegressionSummaryfeatureImportances(classinpyspark.ml.regression),313park.ml.classification.RandomForestClassificationModelGeneralizedLinearRegressionTraining
Summaryattribute),340
(classinpyspark.ml.regression),314featureImportances(pys-getAlpha()(pyspark.ml.recommendation.ALSpark.ml.regression.DecisionTreeRegressionModelmethod),364 attribute),306
getBlockSize() (pys-featureImportances(pys- park.ml.classification.MultilayerPerceptronClassifierpark.ml.regression.GBTRegressionModel method),344
attribute),309 getCensorCol() (pys- featureImportances (pys- park.ml.regression.AFTSurvivalRegression
park.ml.regression.RandomForestRegressionModel method),304attribute),323 getCheckpointFiles() (pysfeaturesCol (pys-
park.ml.clustering.DistributedLDAModel park.ml.classification.LogisticRegressionSummary
method),359attribute),330getColdStartStrategy()(pysfeaturesCol(pys-park.ml.recommendation.ALSmetho), park.ml.regression.LinearRegressionSummary 364 attribute),319
getDistanceMeasure() (pys- fMeasureByLabel() (pys- park.ml.clustering.BisectingKMeans park.ml.classification.LogisticRegressionSummary
method),348method),330 getDistanceMeasure() (pysfMeasureByThreshold (pys- park.ml.clustering.KMeans
method),park.ml.classification.BinaryLogisticRegressionSummary350attribute),332GGaussianMixture (class inpark.ml.clustering),352GaussianMixtureModel(class
inpark.ml.clustering),353GaussianMixtureSummary(class inpark.ml.clustering),354pyspyspys- getDistanceMeasure()
(pyspark.ml.evaluation.ClusteringEvaluatormethod),379getDocConcentration() (pyspark.ml.clustering.LDAMethod),355getDstCol()
(pyspark.ml.clustering.PowerIterationClusteringmethod),362getEpsilon() (pys- gaussiansDF (pyspark.ml.regression.LinearRegression
park.ml.clustering.GaussianMixtureModel method),318 getFamily() (pys-getLearningDecay() (pyspark.ml.classification.LogisticRegression
park.ml.clustering.LDAMethod),356method),327 getLearningOffset() (pysgetFamily() (pyspark.ml.clustering.LDAMethod),356
park.ml.regression.GeneralizedLinearRegressiongetLink() (pys-method),311getFeatureIndex() (pyspark.ml.regression.IsotonicRegressionmethod),316getFinalStorageLevel()
(pys- park.ml.recommendation.ALS method),364getImplicitPrefs() (pyspark.ml.recommendation.ALS method),365getInitialWeights() (pys-
park.ml.regression.GeneralizedLinearRegressionmethod),311getLinkPower() (pyspark.ml.regression.GeneralizedLinearRegressionmethod),311getLinkPredictionCol()
(pyspark.ml.regression.GeneralizedLinearRegressionmethod),311getLossType()
(pys-park.ml.classification.GBTClassifiermethod),337park.ml.classification.MultilayerPerceptronClassifiergetLossType()(pysmethod),344park.ml.regression.GBTRegressormethod),
getInitMode()(pyspark.ml.clustering.KMeans 308 method),350 getLowerBoundsOnCoefficients()(pys- getInitMode()
(pyspark.ml.classification.LogisticRegression park.ml.clustering.PowerIterationClustering method),327method),362getInitSteps() (pys- park.ml.clustering.KMeans
method),350getIntermediateStorageLevel()(pyspark.ml.recommendation.ALSmetho),365getIsotonic()
(pys-park.ml.regression.IsotonicRegressionmethod),316getItemCol()(pyspark.ml.recommendation.ALSmetho),365getK()(pyspark.ml.clustering.BisectingKMeansmethod),348getK()(pyspark
.ml.clustering.GaussianMixturemethod),353getK()(pyspark.ml.clustering.KMeansmethod),351getK()(pyspark.ml.clustering.LDAMethod),355 getLowerBoundsOnIntercepts()
(pys-park.ml.classification.LogisticRegressionmethod),327getMetricName() (pyspark.ml.evaluation.BinaryClassificationEvaluatormethod),376getMetricName()
(pys-park.ml.evaluation.ClusteringEvaluatormethod),379getMetricName() (pys-park.ml.evaluation.MulticlassClassificationEvaluatormethod),378getMetricName()
(pys-park.ml.evaluation.RegressionEvaluatormethod),377getMinDivisibleClusterSize() (pys-park.ml.clustering.BisectingKMeansmethod),348getModelType()
(pys-getK()(pyspark.ml.clustering.PowerIterationClustering park.ml.classification.NaiveBayesmethod),method),362
342getKeepLastCheckpoint()(pys-getNonnegative()(pyspark.ml.clustering.LDAMethod),356park.ml.recommendation.ALSmetho),getLayers()(pys-365park.ml.classification.MultilayerPercep
tronClassifiergetNumFolds()(pysmethod),344park.ml.tuning.CrossValidatormethod),372getNumItemBlocks()park.ml.recommendation.ALS365getNumUserBlocks()park.ml.recommendation.
ALS365getOffsetCol() (pysmethod),(pysmethod),(pys- park.ml.clustering.LDAMethod),356getTopicDistributionCol() (pyspark.ml.clustering.LDAMethod),356getTrainRatio()
(pys-park.ml.tuning.TrainValidationSplitmethod),374getUpperBoundsOnCoefficients()(pys-park.ml.classification.LogisticRegression park.ml.regression.GeneralizedLinearRegression
method),328method),311getOptimizeDocConcentration()(pyspark.ml.clustering.LDAMethod),356getOptimizer()(pyspark.ml.clustering.LDAMethod),356getQuantileProbabilities()
(pys-park.ml.regression.AFTSurvivalRegressionmethod),304getQuantilesCol() (pys-park.ml.regression.AFTSurvivalRegressionmethod),304getRank()
(pyspark.ml.recommendation.ALSmetho),365getRatingCol() (pys- park.ml.recommendation.ALS method),365getSmoothing()
(pyspark.ml.classification.NaiveBayesmethod),342getSrcCol() (pys- getUpperBoundsOnIntercepts()
(pyspark.ml.classification.LogisticRegressionmethod),328getUserCol()(pyspark.ml.recommendation.ALSmetho),365getVariancePower()
(pyspark.ml.regression.GeneralizedLinearRegressionmethod),311HhasSummary
(pys-park.ml.classification.LogisticRegressionModelattribute),329hasSummary
(pyspark.ml.clustering.BisectingKMeansModelattribute),349hasSummary
(pyspark.ml.clustering.GaussianMixtureModelattribute),353hasSummary
(pys-park.ml.clustering.PowerIterationClustering park.ml.clustering.KMeansModel atmethod),362 tribute),351 getStagePath() (pys-hasSummary
(pys-park.ml.pipeline.PipelineSharedReadWrite park.ml.regression.GeneralizedLinearRegressionModelstaticmethod),370 attribute),313getStages()
(pyspark.ml.pipeline.PipelinehasSummary (pysmethod),368 park.ml.regression.LinearRegressionModelgetStepSize() (pys-
attribute),318park.ml.classification.MultilayerPerceptronClassifiermethod),344 getSubsamplingRate() (pyspark.ml.clustering.LDAMethod),356getThreshold()
(pys-park.ml.classification.LogisticRegressionmethod),327getThresholds() (pys-park.ml.classification.LogisticRegressionmethod),328getTopicConcentration() (pys- intercept

(pys-park.ml.classification.LinearSVCModelattribute),325intercept (pys-park.ml.classification.LogisticRegressionModelattribute),329intercept
(pys-park.ml.regression.AFTSurvivalRegressionModelattribute),3051 intercept
(pys-LinearRegressionTrainingSummarypark.ml.regression.GeneralizedLinearRegressionModel(classin pyspark.ml.regression),321attribute),313
LinearSVC(classin pyspark.ml.classification),intercept (pys- 324park.ml.regression.LinearRegressionModelLinearSVCModel(classin pysattribute),318park.ml.classification),325
interceptVector (pys-load()(pyspark.ml.pipeline.PipelineModelReader park.ml.classification.LogisticRegressionModel method),369attribute),329isDistributed()
(pys- park.ml.clustering.LDAModel method),358isLargerBetter() (pys-park.ml.evaluation.Evaluator method),376IsotonicRegression (class in
park.ml.regression),315IsotonicRegressionModel(class in pyspark.ml.regression),316ItemFactors
(pys-park.ml.recommendation.ALSCModelattribute),367KKMeans(classin pyspark.ml.clustering),349KMeansModel(classin pyspark.ml.clustering),351KolmogorovSmirnovTest(class in
pyspark.ml.stat),299L load()
(pyspark.ml.pipeline.PipelineReadermethod),370load()(pyspark.ml.pipeline.PipelineSharedReadWritestaticmethod),370LocalLDAModel(classin pyspark.ml.clustering),359LogisticRegression
(class in pyspark.ml.classification),325LogisticRegressionModel(class in
pyspark.ml.classification),329LogisticRegressionSummary(classin pys-park.ml.classification),330LogisticRegressionTrainingSummary(classin pyspark.ml.classification),331logLikelihood
(pys-park.ml.clustering.GaussianMixtureSummaryattribute),354logLikelihood() (pys- park.ml.clustering.LDAModel method),359logPerplexity() (pys-
park.ml.clustering.LDAModel method),359labelCol(pyspark.ml.classification.LogisticRegressionSummarylogPrior() (pys- attribute),330
park.ml.clustering.DistributedLDAModellabelCol(pyspark.ml.regression.LinearRegressionSummarymethod),360attribute),319labels(pyspark.ml.classification.LogisticRegressionSummaryM
attribute),330 max()(pyspark.ml.stat.Summarizerstaticmethod),layers(pyspark.ml.classification.MultilayerPerceptronClassificationModel301attribute),345
mean() (pyspark.ml.stat.Summarizer staticLDA(classin pyspark.ml.clustering),354 method),301 LDAModel(classin pyspark.ml.clustering),358
meanAbsoluteError (pys-LinearRegression (class in park.ml.regression),316LinearRegressionModel(class in park.ml.regression),318LinearRegressionSummary(class
in park.ml.regression),319 pyspyspys- park.ml.regression.LinearRegressionSummaryattribute),320meanSquaredError
(pys-park.ml.regression.LinearRegressionSummaryattribute),320metrics()(pyspark.ml.stat.Summarizer
staticmethod),301min()(pyspark.ml.stat.Summarizerstaticmethod),pi(pyspark.ml.classification.NaiveBayesModelat- 302 tribute),342 MulticlassClassificationEvaluator
Pipeline(classin pyspark.ml.pipeline),368 (classin pyspark.ml.evaluation),377 PipelineModel(classin pyspark.ml.pipeline), MultilayerPerceptronClassificationModel
369(classin pyspark.ml.classification),345MultilayerPerceptronClassifier(classin pyspark.ml.classification),343NNaiveBayes(classin pyspark.ml.classification),341NaiveBayesModel
(class in pyspark.ml.classification),342normL1()(pyspark.ml.stat.Summarizer staticmethod),302normL2()(pyspark.ml.stat.Summarizer
staticmethod),302nullDeviance (pys- PipelineModelReader(class in pyspark.ml.pipeline),369PipelineModelWriter(class in
pyspark.ml.pipeline),369PipelineReader(classin pyspark.ml.pipeline),369PipelineSharedReadWrite(class in
pyspark.ml.pipeline),370PipelineWriter(classin pyspark.ml.pipeline),370PowerIterationClustering(classin pys-park.ml.clustering),360pr(pyspark.ml.classification.BinaryLogisticRegressionSum
maryattribute),332 park.ml.regression.GeneralizedLinearRegressionSummaryprecisionByLabel
(pys-attribute),313park.ml.classification.LogisticRegressionSummarynumInstances(pys-attribute),330
park.ml.regression.GeneralizedLinearRegressionSummaryprecisionByThreshold
(pys-attribute),314park.ml.classification.BinaryLogisticRegressionSummarynumInstances(pys-attribute),332park.ml.regression.LinearRegressionSummarypredict()(pysattribute),320park.ml.r
egression.AFTSurvivalRegressionModelnumIterations(pys-method),305 park.ml.regression.GeneralizedLinearRegressionTrainingSummarypredictionCol (pys- attribute),315
park.ml.classification.LogisticRegressionSummary numNonZeros()(pyspark.ml.stat.Summarizer attribute),330 staticmethod),302 predictionCol
(pys-park.ml.regression.GeneralizedLinearRegressionSummary O attribute),314 objectiveHistory (pys-predictionCol
(pys-park.ml.classification.LogisticRegressionTrainingSummarypark.ml.regression.LinearRegressionSummaryattribute),331 attribute),320 objectiveHistory
(pys-predictions (pys-park.ml.regression.LinearRegressionTrainingSummarypark.ml.classification.LogisticRegressionSummary attribute),322 attribute),330
OneVsRest(classin pyspark.ml.classification),predictions (pys-345OneVsRestModel (classspark.ml.classification),347PPParamGridBuilder (classspark.ml.tuning),370
inin pys-pys- park.ml.regression.GeneralizedLinearRegressionSummaryattribute),314predictions
(pys-park.ml.regression.IsotonicRegressionModelattribute),316predictions
(pys-park.ml.regression.LinearRegressionSummaryattribute),320read()(pyspark.ml.tuning.CrossValidatorclasspredictQuantiles()(pys-method),372park.ml.regression.AFTSurvivalRegression
Modelread()(pyspark.ml.tuning.CrossValidatorModelmethod),305 classmethod),373 probability
(pys-read()(pyspark.ml.tuning.TrainValidationSplitpark.ml.clustering.GaussianMixtureSummaryclassmethod),374attribute),354read()(pyspark.ml.tuning.TrainValidationSplitModelprobabilityC
ol(pys-classmethod),375park.ml.classification.LogisticRegressionSummaryrecallByLabel(pysattribute),331 park.ml.classification.LogisticRegressionSummaryprobabilityCol(pys-attribute),331
park.ml.clustering.GaussianMixtureSummaryrecallByThreshold(pysattribute),354park.ml.classification.BinaryLogisticRegressionSummarypValues(pyspark.ml.regression.GeneralizedLinearR
egressionTraattributionSummary),333attribute),315 recommendForAllItems() (pyspValues(pyspark.ml.regression.LinearRegressionSummary
park.ml.recommendation.ALSCModelattribute),320pyspark.ml.classification(module),324pyspark.ml.clustering(module),347pyspark.ml.evaluation(module),375pyspark.ml.pipeline(module),36
8pyspark.ml.recommendation(module),363pyspark.ml.regression(module),303pyspark.ml.stat(module),297pyspark.ml.tuning(module),370R
method),367recommendForAllUsers()(park.ml.recommendation.ALSCModelmethod),367recommendForItemSubset()(park.ml.recommendation.ALSCModelmethod),367recommendForUserSubs
et()(park.ml.recommendation.ALSCModelmethod),367RegressionEvaluator(class in (pys-(pys-(pys-pys- r2(pyspark.ml.regression.LinearRegressionSummary
park.ml.evaluation),377 attribute),320 residualDegreeOfFreedom
(pys-r2adj(pyspark.ml.regression.LinearRegressionSummarypark.ml.regression.GeneralizedLinearRegressionSummaryattribute),321 attribute),314
RandomForestClassificationModel residualDegreeOfFreedomNull (pys- (classin pyspark.ml.classification),340park.ml.regression.GeneralizedLinearRegressionSummary
RandomForestClassifier(class in pys- attribute),314 park.ml.classification),338 residuals
(pys-RandomForestRegressionModel(classin park.ml.regression.LinearRegressionSummarypyspark.ml.regression),323attribute),321 RandomForestRegressor(class in
pys-residuals() (pys- park.ml.regression),322
park.ml.regression.GeneralizedLinearRegressionSummaryrank(pyspark.ml.recommendation.ALSCModelat-method),314tribute),367roc(pyspark.ml.classification.BinaryLogisticRegressionSu
mmaryrank(pyspark.ml.regression.GeneralizedLinearRegressionSummaryattribute),333 attribute),314 rootMeanSquaredError
(pys-read()(pyspark.ml.pipeline.Pipelineclasspark.ml.regression.LinearRegressionSummarymethod),368
attribute),321read()(pyspark.ml.pipeline.PipelineModelclassRunonDatabricksCommunityCloud,method),369 11 S method),312 saveImpl()
(pys-setFeatureIndex() (pys-park.ml.pipeline.PipelineModelWriter park.ml.regression.IsotonicRegressionmethod),369 method),316 saveImpl()
(pys-setFeatureSubsetStrategy() (pys-park.ml.pipeline.PipelineSharedReadWrite park.ml.classification.GBTClassifierstaticmethod),370 method),337
saveImpl()(pyspark.ml.pipeline.PipelineWritersetFeatureSubsetStrategy() (pys- method),370 park.ml.classification.RandomForestClassifier
scale(pyspark.ml.regression.AFTSurvivalRegressionModel
method),340attribute),305setFeatureSubsetStrategy()(pysyscale(pyspark.ml.regression.LinearRegressionModelpark.ml.regression.GBTRegressormethod),attribute),318
308SetupSparkonCloud,20setFeatureSubsetStrategy()(pyssetAlpha()(pyspark.ml.recommendation.ALSPark.ml.regression.RandomForestRegressormethod),365 method),323setBlockSize()
(pys-setFinalStorageLevel() (pyspark.ml.classification.MultilayerPerceptronClassifierpark.ml.recommendation.ALS method),method),344setCensorCol()
(pys-park.ml.regression.AFTSurvivalRegressionmethod),304setColdStartStrategy() (pys- park.ml.recommendation.ALS method),365setDistanceMeasure()
(pys-park.ml.clustering.BisectingKMeansmethod),348setDistanceMeasure() (pys- park.ml.clustering.KMeans method),351setDistanceMeasure()
(pys-park.ml.evaluation.ClusteringEvaluatormethod),379setDocConcentration() (pyspark.ml.clustering.LDAMethod),356setDstCol() (pys- 365setImplicitPrefs()
(pys- park.ml.recommendation.ALS method),366setInitialWeights()
(pys-park.ml.classification.MultilayerPerceptronClassifiermethod),345setInitMode()(pyspark.ml.clustering.KMeansmethod),351setInitMode()
(pys-park.ml.clustering.PowerIterationClusteringmethod),362setInitSteps() (pys- park.ml.clustering.KMeans method),351setIntermediateStorageLevel()(pys-
park.ml.recommendation.ALS method),366setIsotonic() (pys-park.ml.clustering.PowerIterationClustering park.ml.regression.IsotonicRegressionmethod),362
method),316 setEpsilon() (pys-setItemCol()(pyspark.ml.recommendation.ALS park.ml.regression.LinearRegression
method),366method),318setK()(pyspark.ml.clustering.BisectingKMeanssetFamily()(pys-method),348park.ml.classification.LogisticRegressionsetK()(pyspark.ml.clustering.GaussianMixture
method),328 method),353 setFamily() (pys-setK()(pyspark.ml.clustering.KMeansmethod), park.ml.regression.GeneralizedLinearRegression
351setK()(pyspark.ml.clustering.LDAMethod),356setK()(pyspark.ml.clustering.PowerIterationClustering park.ml.classification.NaiveBayesmethod), method),362
342setKeepLastCheckpoint() (pys-setNonnegative()park.ml.clustering.LDAMethod),357 park.ml.recommendation.ALSsetLayers() (pys-
366park.ml.classification.MultilayerPerceptronClassifiersetNumBlocks() method),345 park.ml.recommendation.ALSsetLearningDecay()(pys- 366
park.ml.clustering.LDAMethod),357 setNumFolds()(setLearningOffset() (pys- park.ml.tuning.CrossValidatorpark.ml.clustering.LDAMethod),357 372setLink()

(pys-setNumItemBlocks())park.ml.regression.GeneralizedLinearRegressionpark.ml.recommendation.ALSSmethod),312 366setLinkPower()
(pys-setNumUserBlocks())park.ml.regression.GeneralizedLinearRegression park.ml.recommendation.ALSSmethod),312 366setLinkPredictionCol()
(pys-setOffsetCol() (pysmethod),(pysmethod),(pysmethod),(pysmethod),(pys- park.ml.regression.GeneralizedLinearRegression
park.ml.regression.GeneralizedLinearRegressionmethod),312setLossType() (pys-park.ml.classification.GBTClassifiermethod),337setLossType()
(pys-park.ml.regression.GBTRegressormethod),308setLowerBoundsOnCoefficients()(pys-park.ml.classification.LogisticRegressionmethod),328setLowerBoundsOnIntercepts()
(pys-park.ml.classification.LogisticRegressionmethod),328setMetricName() (pys-
method),312setOptimizeDocConcentration()(pyspark.ml.clustering.LDAMethod),357setOptimizer()(pyspark.ml.clustering.LDAMethod),357setParams()
(pys-park.ml.classification.DecisionTreeClassifiermethod),335setParams() (pys-park.ml.classification.GBTClassifiermethod),337setParams()
(pys-park.ml.classification.LinearSVCmethod),325park.ml.evaluation.BinaryClassificationEvaluatorsetParams()(pysmethod),376park.ml.classification.LogisticRegressionsetMetricName()(py
s-method),328park.ml.evaluation.ClusteringEvaluatorsetParams()(pysmethod),379park.ml.classification.MultilayerPerceptronClassifiersetMetricName()(pys-method),345park.ml.evaluation.
MulticlassClassificationEvaluatorsetParams()(pysmethod),378park.ml.classification.NaiveBayesmethod),setMetricName()(pys-342park.ml.evaluation.RegressionEvaluatorsetParams()(pysm
ethod),377park.ml.classification.OneVsRestmethod),setMinDivisibleClusterSize()(pys-346park.ml.clustering.BisectingKMeanssetParams()(pysmethod),349park.ml.classification.RandomFor
estClassifiersetModelType() (pys- method),340 setParams()(pys-
park.ml.regression.RandomForestRegressorpark.ml.clustering.BisectingKMeansmethod),323method),349setParams()(pyspark.ml.tuning.CrossValidatorsetParams()(pys-method),372park.m
l.clustering.GaussianMixturesetParams()(pysmethod),353park.ml.tuning.TrainValidationSplit setParams()(pyspark.ml.clustering.KMeans method),374 method),351
setQuantileProbabilities() (pys- setParams() (pyspark.ml.clustering.LDA park.ml.regression.AFTSurvivalRegressionmethod),357 method),304
setParams()(pys-setQuantilesCol() (pys-park.ml.clustering.PowerIterationClustering park.ml.regression.AFTSurvivalRegressionmethod),362
method),304setParams()(pys-setRank()(pyspark.ml.recommendation.ALSPark.ml.evaluation.BinaryClassificationEvaluatormethod),366method),377 setRatingCol()
(pyssetParams() (pys- park.ml.recommendation.ALS method), park.ml.evaluation.ClusteringEvaluator
366method),379setSmoothing()(pyssetParams()(pys-park.ml.classification.NaiveBayesmethod),park.ml.evaluation.MulticlassClassificationEvaluator342method),378setParams()
(pys-park.ml.evaluation.RegressionEvaluatormethod),377setParams() (pyspark.ml.pipeline.Pipelinemethod),369setParams()(pyspark.ml.recommendation.ALSSmethod),366setParams()
(pys-park.ml.regression.AFTSurvivalRegressionmethod),304setParams() (pys-park.ml.regression.DecisionTreeRegressormethod),306setParams()
(pys-park.ml.regression.GBTRegressormethod),308setParams() (pys- setSrcCol() (pys-park.ml.clustering.PowerIterationClusteringmethod),362setStages()
(pyspark.ml.pipeline.Pipelinemethod),369setStepSize() (pys-park.ml.classification.MultilayerPerceptronClassifiermethod),345setSubsamplingRate()
(pyspark.ml.clustering.LDAMethod),358setThreshold() (pys-park.ml.classification.LogisticRegressionmethod),328setThresholds()
(pys-park.ml.classification.LogisticRegressionmethod),328setTopicConcentration()
(pys-park.ml.clustering.LDAMethod),358park.ml.regression.GeneralizedLinearRegressionsetTopicDistributionCol()(pysmethod),312park.ml.clustering.LDAMethod),358 setParams()
(pys-setTrainRatio() (pys-park.ml.regression.IsotonicRegression park.ml.tuning.TrainValidationSplitmethod),316 method),374 setParams()
(pys-setUpperBoundsOnCoefficients()(pys-park.ml.regression.LinearRegression park.ml.classification.LogisticRegressionmethod),318 method),329setParams()
(pys-setUpperBoundsOnIntercepts() (pyspark.ml.classification.LogisticRegressiontotalIterations(pysmethod),329park.ml.regression.LinearRegressionTrainingSummary
setUserCol()(pyspark.ml.recommendation.ALS attribute),322 method),366 trainingLogLikelihood() (pys- setVariancePower() (pys-
park.ml.clustering.DistributedLDAModel park.ml.regression.GeneralizedLinearRegression method),360 method),312 TrainValidationSplit(class in

pys-solver(pyspark.ml.regression.GeneralizedLinearRegressionTrainingSummarypark.ml.tuning),373attribute),315TrainValidationSplitModel(classinpyssubModels(pys-park.ml.tuning),374pa
rk.ml.tuning.CrossValidatorModel trees(pyspark.ml.classification.GBTClassificationModelattribute),373 attribute),338 subModels
(pys-trees(pyspark.ml.classification.RandomForestClassificationModel park.ml.tuning.TrainValidationSplitModel attribute),341attribute),375
trees(pyspark.ml.regression.GBTRegressionModelSummarizer(classinpysspark.ml.stat),300
attribute),309summary(pyspark.ml.classification.LogisticRegressionModeltrees(pyspark.ml.regression.RandomForestRegressionModelattribute),329
attribute),324summary(pyspark.ml.clustering.BisectingKMeansModeltruePositiveRateByLabel(pysattribute),349park.ml.classification.LogisticRegressionSummarysummary(pyspark.ml.cluste
ring.GaussianMixtureModelattribute),331attribute),354tValues(pyspark.ml.regression.GeneralizedLinearRegressionTrainsummary(pyspark.ml.clustering.KMeansModelat-attribute),315tribute
,352tValues(pyspark.ml.regression.LinearRegressionSummarysummary(pyspark.ml.regression.GeneralizedLinearRegressionModelattribute),321attribute),313summary(pyspark.ml.regressi
on.LinearRegressionModelUattribute),318summary()
(pyspark.ml.stat.SummaryBuildermethod),302SummaryBuilder(classinpysspark.ml.stat),302Ttest()(pyspark.ml.stat.ChiSquareTest
staticmethod),297test()(pyspark.ml.stat.KolmogorovSmirnovTeststaticmethod),299theta(pyspark.ml.classification.NaiveBayesModelattribute),342toLocal()
(pys-park.ml.clustering.DistributedLDAModelmethod),360topicsMatrix() (pys- park.ml.clustering.LDAModel method),359totalIterations (pys- userFactors
(pys-park.ml.recommendation.ALSModelattribute),368VvalidateStages() (pys-park.ml.pipeline.PipelineSharedReadWritestaticmethod),370validationMetrics
(pys-park.ml.tuning.TrainValidationSplitModelattribute),375variance()(pyspark.ml.stat.Summarizerstaticmethod),302vocabSize()(pyspark.ml.clustering.LDAModelmethod),359WweightedFals
ePositiveRate (pys-park.ml.classification.LogisticRegressionSummarypark.ml.classification.LogisticRegressionTrainingSummaryattribute),331attribute),332
weightedFMeasure() (pys-park.ml.classification.LogisticRegressionSummarymethod),331 weightedPrecision
(pys-park.ml.classification.LogisticRegressionSummaryattribute),331 weightedRecall (pys-park.ml.classification.LogisticRegressionSummaryattribute),331
weightedTruePositiveRate
(pys-park.ml.classification.LogisticRegressionSummaryattribute),331weights(pyspark.ml.classification.MultilayerPerceptronClassificationModelattribute),345weights(pyspark.ml.clustering.Ga
ussianMixtureModelattribute),354write()(pyspark.ml.pipeline.Pipelinemethod),369write() (pyspark.ml.pipeline.PipelineModelmethod),369write()
(pyspark.ml.tuning.CrossValidatormethod),372write()(pyspark.ml.tuning.CrossValidatorModelmethod),373write()(pyspark.ml.tuning.TrainValidationSplitmethod),374write()(pyspark.ml.tuning.
TrainValidationSplitMethod),375Viewpublicationstats

`park.ml.regression.GeneralizedLinearRegressionModel`/`DecisionTreeClassifier`(class in `pyspark.ml.classification`),333
coefficients (`pyspark.ml.regression.LinearRegressionModel` `pyspark.ml.regression`),306
attribute),318 `DecisionTreeRegressor`(class in `pyspark.ml.regression`),305
coefficientStandardErrors (`pyspark.ml.regression.GeneralizedLinearRegressionTrainingSummary`degreesOfFreedom
attribute),314 `park.ml.regression.GeneralizedLinearRegressionSummary`
attribute),319 `park.ml.regression.LinearRegressionSummary`attribute),319
computeCost() (`pyspark.ml.clustering.BisectingKMeansModel`method),349 `park.ml.clustering.LDAModel` method),
computeCost() (`pyspark.ml.clustering.KMeansModel`method),351 `deviance`(`pyspark.ml.regression.GeneralizedLinearRegressionSum`attribute),313
ConfigureSparkonMacandUbuntu, `devianceResiduals` (`pyspark.ml.regression.LinearRegressionSummary`attribute),319
16copy() (`pyspark.ml.classification.OneVsRestModel`method),346 `dispersion` (`pyspark.ml.regression.GeneralizedLinearRegressionSummary`attribute),313
copy()(`pyspark.ml.classification.OneVsRestModel`method),347 `DistributedLDAModel`(class in `pyspark.ml.clustering`),359
copy()(`pyspark.ml.pipeline.Pipeline` method), `E`
368copy() (`pyspark.ml.pipeline.PipelineModel`method),369copy() `estimatedDocConcentration`() (`pyspark.ml.tuning.CrossValidator`method),372
(`pyspark.ml.tuning.CrossValidatorModel`method),373 `park.ml.clustering.LDAModel` method),
copy()(`pyspark.ml.tuning.TrainValidationSplit` method),374 `evaluate`() (`pyspark.ml.classification.LogisticRegressionModel`method),329
method),374

coefficientStandardErrors(`pyspark.ml.regression.LinearRegressionSummary`degreesOfFreedom(`pyspark.ml.tuning.TrainValidationSplitModel`evaluate())(`pyspark.ml.evaluation.Evaluator`
method),375
corr() (`pyspark.ml.stat.Correlation` staticmethod),298
Correlation(classin`pyspark.ml.stat`),298count()(`pyspark.ml.stat.Summarizer`
staticmethod),301
CrossValidator(classin`pyspark.ml.tuning`),371
CrossValidatorModel(class in `pyspark.ml.tuning`),372

```

explainedVariance (pys-
park.ml.regression.LinearRegressionSummaryGBTClassificationMo
F
falsePositiveRateByLabel(pys-GBTCRegressionModel(classinpyspark.ml.classifica
attribute),330 GBTCRegressor(classinpyspark.ml.regression),featu
park.ml.classification.DecisionTreeClassificationModelGeneralizedL
featureImportances
park.ml.classification.GBTCClassficationModel
attribute),338
featureImportances (pys- (classinpyspark.ml.regression),313park
(classinpyspark.ml.regression),314
featureImportances(pys-getAlpha()(pyspark.ml.recommendation.ALSpark.ml.regr
attribute),306
featureImportances (pys- park.ml.classification.MultilayerPercepti
attribute),309
featureImportances (pys- (classinpyspark.ml.regression.RandomForestRegressionModel
attribute),323 getCheckpointFiles() (pysfeaturesCol (p
park.ml.classification.LogisticRegressionSummary
attribute),330getColdStartStrategy()(pysfeaturesCol(pys-park.ml.re
park.ml.regression.LinearRegressionSummary
attribute),319
f1MeasureByLabel() (pys-
park.ml.classification.LogisticRegressionSummary
method),330 getDistanceMeasure() (pysf1MeasureByThreshold
park.ml.classification.BinaryLogisticRegressionSummary350

```

attribute),332

G

GaussianMixture (class in park.ml.clustering),352

GaussianMixtureModel(class in park.ml.clustering),353

GaussianMixtureSummary(class in park.ml.clustering),354

getDistanceMeasure() (pyspark.ml.evaluation.ClusteringEvaluator),379

getDocConcentration() (pyspark.ml.clustering.LDAmethod),355

getDstCol() (pyspark.ml.clustering.PowerIterationClusteringmethod),362

getEpsilon() (pyspark.ml.clustering.PowerIterationClusteringmethod),362

gaussiansDF (pyspark.ml.clustering.GaussianMixtureModel)

getFamily() (pyspark.ml.classification.LogisticRegressionmethod),327

getLearningOffset() (pyspark.ml.clustering.LDAmethod),356

getLearningOffset() (pyspark.ml.clustering.LDAmethod),356

getLink() (pyspark.ml.regression.GeneralizedLinearRegressionmethod),311

getFeatureIndex() (pyspark.ml.regression.IsotonicRegressionmethod),316

getFinalStorageLevel() (pyspark.ml.recommendation.ALSmethod),364

getImplicitPrefs() (pyspark.ml.recommendation.ALSmethod),365

getInitialWeights() (pyspark.ml.recommendation.ALSmethod),365

getLossType() (pyspark.ml.classification.GBTClassifiermethod),337

getLossType() (pyspark.ml.regression.GBTRegressormethod),344

getLowerBoundsOnCoefficients() (pyspark.ml.classification.LogisticRegressionmethod),327

getLowerBoundsOnIntercepts() (pyspark.ml.classification.LogisticRegressionmethod),327

getMetricName() (pyspark.ml.evaluation.BinaryClassificationEvaluator),376

getMetricName() (pyspark.ml.evaluation.ClusteringEvaluator),379

getMetricName() (pyspark.ml.evaluation.MulticlassClassificationEvaluator),378

getMetricName() (pyspark.ml.evaluation.RegressionEvaluator),377

getMinDivisibleClusterSize() (pyspark.ml.clustering.BisectingKMeansmethod),348

getModelType() (pyspark.ml.classification.NaiveBayesmethod),342

getKeepLastCheckpoint() (pyspark.ml.clustering.LDAmethod),356

getKeepLastCheckpoint() (pyspark.ml.recommendation.ALSmethod),365

getKeepLastCheckpoint() (pyspark.ml.classification.MultilayerPerceptronClassifiergetNumFolds()),344

getKeepLastCheckpoint() (pyspark.ml.tuning.CrossValidator),342

Index

attribute),329
 isDistributed()
 park.ml.clustering.LDAModel
 358
 isLargerBetter()
 park.ml.evaluation.Evaluator method),376
 IsotonicRegression (class in pyspark.ml.regression),315
 IsotonicRegressionModel(class in pyspark.ml.regression),316
 itemFactors
 park.ml.recommendation.ALSModelattribute),367
 K
 KMeans(classinpyspark.ml.clustering),349
 KMeansModel(classinpyspark.ml.clustering),351
 KolmogorovSmirnovTest(class in pyspark.ml.stat),299
 L
 load() (pyspark.ml.pipeline.PipelineReadermethod),370
 (pys-load()(pyspark.ml.pipeline.PipelineSharedReadWritestaticmethod),370
 method),LocalLDAModel(classinpyspark.ml.clustering),359
 LogisticRegression (class in pyspark.ml.classification),325
 (pys-LogisticRegressionModel(class in pyspark.ml.classification),329
 LogisticRegressionSummary(classinpys-
 park.ml.classification),330
 LogisticRegressionTrainingSummary
 (classinpyspark.ml.classification),331
 logLikelihood (pys-
 park.ml.clustering.GaussianMixtureSummaryattribute),354
 logLikelihood() (pys-
 park.ml.clustering.LDAModel method),
 359
 logPerplexity() (pys-
 park.ml.clustering.LDAModel method),
 359
 labelCol(pyspark.ml.classification.LogisticRegressionSummarylogPrior() (pys-
 attribute),330
 park.ml.clustering.DistributedLDAModel
 labelCol(pyspark.ml.regression.LinearRegressionSummarymethod),360attribute),319
 labels(pyspark.ml.classification.LogisticRegressionSummaryM
 attribute),330
 max()(pyspark.ml.stat.Summarizerstaticmethod),
 layers(pyspark.ml.classification.MultilayerPerceptronClassificationModel301
 attribute),345 mean() (pyspark.ml.stat.Summarizer staticLDA(classinpyspark.ml.clustering),354 method),301
 LDAModel(classinpyspark.ml.clustering),358
 meanAbsoluteError (pys-
 LinearRegression (classpyspyspys- park.ml.regression.LinearRegressionSummaryattribute),320
 in meanSquaredError (pys-
 park.ml.regression),316
 park.ml.regression.LinearRegressionSummaryattribute),320metrics()(pyspark.ml.stat.Summarizer
 LinearRegressionModel(class in static
 park.ml.regression),318 method),301
 LinearRegressionSummary(class in
 park.ml.regression),319
 min()(pyspark.ml.stat.Summarizerstaticmethod),pi(pyspark.ml.classification.NaiveBayesModelat-
 302
 MulticlassClassificationEvaluator
 (classinpyspark.ml.evaluation),377
 MultilayerPerceptronClassificationModel
 tribute),342
 Pipeline(classinpyspark.ml.pipeline),368
 PipelineModel(classinpyspark.ml.pipeline),
 369

(classin `pyspark.ml.classification`), 345
 MultilayerPerceptronClassifier(classin `pyspark.ml.classification`), 343
 N
 NaiveBayes(classin `pyspark.ml.classification`), 341
 NaiveBayesModel (class in `pyspark.ml.classification`), 342
 normL1()(`pyspark.ml.stat.Summarizer` staticmethod), 302
 normL2()(`pyspark.ml.stat.Summarizer` staticmethod), 302
 nullDeviance

 PipelineModelReader(class in `pyspark.ml.pipeline`), 369
 PipelineModelWriter(class in `pyspark.ml.pipeline`), 369
 PipelineReader(classin `pyspark.ml.pipeline`), 369
 PipelineSharedReadWrite(class in `pyspark.ml.pipeline`), 370
 PipelineWriter(classin `pyspark.ml.pipeline`), 370
 PowerIterationClustering(classin `pyspark.ml.clustering`), 360
 pr(`pyspark.ml.classification.BinaryLogisticRegressionSummary` attribute), 332

`park.ml.regression.GeneralizedLinearRegressionSummary`precisionByLabel attribute), 313
`park.ml.classification.LogisticRegressionSummary`numInstances(`pys-attribute`), 330
`park.ml.regression.GeneralizedLinearRegressionSummary`precisionByThreshold attribute), 314
`park.ml.classification.BinaryLogisticRegressionSummary`numInstances(`pys-attribute`), 332
`park.ml.regression.LinearRegressionSummary`predict()(`pysattribute`), 320
`park.ml.regression.AFTSurvivalRegressionModel`numIterations(`pys-method`), 305
`park.ml.regression.GeneralizedLinearRegressionTrainingSummary`predictionCol attribute), 315
 numNonZeros()(`pyspark.ml.stat.Summarizer` staticmethod), 302
 predictionCol
`park.ml.classification.LogisticRegressionSummary` attribute), 330
`park.ml.regression.GeneralizedLinearRegressionSummary` attribute), 314
 objectiveHistory
`park.ml.classification.LogisticRegressionTrainingSummary``park.ml.regression.LinearRegressionSummary`attribute), 331
 objectiveHistory
`park.ml.regression.LinearRegressionTrainingSummary``park.ml.classification.LogisticRegressionSummary` attribute), 322
 OneVsRest(classin `pyspark.ml.classification`), predictions
 OneVsRestModel (class `park.ml.classification`), 347
 P
 ParamGridBuilder (class in `pyspark.ml.tuning`), 370
 predictions
`park.ml.regression.GeneralizedLinearRegressionSummary`attribute), 314
`park.ml.regression.IsotonicRegressionModel`attribute), 316
 predictions
`park.ml.regression.LinearRegressionSummary`
 attribute), 320
 read()(`pyspark.ml.tuning.CrossValidator`classpredictQuantiles()(`pys-method`), 372
`park.ml.regression.AFTSurvivalRegressionModel`read()(`pyspark.ml.tuning.CrossValidatorModel`method), 305
 classmethod), 373
 probability
`park.ml.clustering.GaussianMixtureSummary`classmethod), 374
 attribute), 354
 read()(`pyspark.ml.tuning.TrainValidationSplitModel`probabilityCol(`pys-classmethod`), 375
`park.ml.classification.LogisticRegressionSummary`recallByLabel(`pysattribute`), 331
`park.ml.classification.LogisticRegressionSummary`probabilityCol(`pys-attribute`), 331
`park.ml.clustering.GaussianMixtureSummary`recallByThreshold(`pysattribute`), 354
`park.ml.classification.BinaryLogisticRegressionSummary`
 pValues(`pyspark.ml.regression.GeneralizedLinearRegressionTraa`attribute
`ningsSummary`), 333
 attribute), 315
 recommendForAllItems() (`pyspValues`(`pyspark.ml.regression.LinearRegressionSummary` `park.ml.recommendation.ALSModel` method), 367
 attribute), 320
`pyspark.ml.classification`(`module`), 324
`pyspark.ml.clustering`(`module`), 347
`pyspark.ml.evaluation`(`module`), 375
`pyspark.ml.pipeline`(`module`), 368
`pyspark.ml.recommendation`(`module`), 363
`pyspark.ml.regression`(`module`), 303
`pyspark.ml.stat`(`module`), 297
`pyspark.ml.tuning`(`module`), 370
 recommendForAllUsers()(`park.ml.recommendation.ALSModel`method), 367
 recommendForItemSubset()(`park.ml.recommendation.ALSModel`method), 367
 recommendForUserSubset()(`park.ml.recommendation.ALSModel`method), 367
 RegressionEvaluator(class in `pyspark.ml.evaluation`), 377
 r2(`pyspark.ml.regression.LinearRegressionSummary` `park.ml.evaluation`), 377

attribute),320 residualDegreeOfFreedom (pys-
r2adj(pyspark.ml.regression.LinearRegressionSummarypark.ml.regression.GeneralizedLinearRegressionSummaryattribute),321 attribute),314
RandomForestClassificationModel residualDegreeOfFreedomNull (pys-
(classin pyspark.ml.classification),340 park.ml.regression.GeneralizedLinearRegressionSummary
RandomForestClassifier(class in pys- attribute),314
park.ml.classification),338 residuals (pys-
RandomForestRegressionModel(classinpark.ml.regression.LinearRegressionSummarypyspark.ml.regression),323attribute),321
RandomForestRegressor(class in pys-residuals()) (pys-
park.ml.regression),322 park.ml.regression.GeneralizedLinearRegressionSummary
rank(pyspark.ml.recommendation.ALSModelat-method),314tribute),367roc(pyspark.ml.classification.BinaryLogisticRegressionSummary
rank(pyspark.ml.regression.GeneralizedLinearRegressionSummaryattribute),333
attribute),314 rootMeanSquaredError (pys-
read()(pyspark.ml.pipeline.Pipelineclasspark.ml.regression.LinearRegressionSummarymethod),368 attribute),321
read()(pyspark.ml.pipeline.PipelineModel/classRunonDatabricksCommunityCloud,method),369 11
S method),312
saveImpl() (pys-setFeatureIndex()) (pys-
park.ml.pipeline.PipelineModelWriter park.ml.regression.IsotonicRegressionmethod),369 method),316
saveImpl() (pys-setFeatureSubsetStrategy()) (pys-
park.ml.pipeline.PipelineSharedReadWrite park.ml.classification.GBTClassifierstaticmethod),370 method),337
saveImpl()(pyspark.ml.pipeline.PipelineWritersetFeatureSubsetStrategy()) (pys-
method),370 park.ml.classification.RandomForestClassifier
scale(pyspark.ml.regression.AFTSurvivalRegressionModel method),340
attribute),305setFeatureSubsetStrategy()(pysscale(pyspark.ml.regression.LinearRegressionModelpark.ml.regression.GBTRegressormethod),attribute),318 308
SetupSparkonCloud,20setFeatureSubsetStrategy()(pyssetAlpha()(pyspark.ml.recommendation.AL Spark.ml.regression.RandomForestRegressormethod),365
method),323
setBlockSize() (pys-setFinalStorageLevel()) (pyspark.ml.classification.MultilayerPerceptronClassifierpark.ml.recommendation.ALS method),
method),344 365
setCensorCol() (pys-setImplicitPrefs()) (pys-
park.ml.regression.AFTSurvivalRegressionmethod),3 04 park.ml.recommendation.ALS method),
setColdStartStrategy() (pys-setInitialWeights()) (pys-
park.ml.recommendation.ALS method), 366 park.ml.classification.MultilayerPerceptronClassifiermethod),345setinitMode()(pyspark.ml.clustering.KMeansmetho
365 d),351
setDistanceMeasure() (pys-setInitMode()) (pys-
park.ml.clustering.BisectingKMeansmethod),348 park.ml.clustering.PowerIterationClusteringmethod),362
setDistanceMeasure() (pys-setInitSteps()) (pys-
park.ml.clustering.KMeans method), 351 park.ml.clustering.KMeans method),
351 351setIntermediateStorageLevel()(pys-
setDistanceMeasure() (pys- park.ml.recommendation.ALS method),
park.ml.evaluation.ClusteringEvaluatormethod),379 366
setDocConcentration() (pyspark.ml.clustering.LDAmethod),356 setIsotonic() (pys-
setDstCol() (pys-
park.ml.clustering.PowerIterationClustering park.ml.regression.IsotonicRegressionmethod),362 method),316
setEpsilon() (pys-setItemCol()(pyspark.ml.recommendation.ALS
park.ml.regression.LinearRegression method),366
method),318setK()(pyspark.ml.clustering.BisectingKMeanssetFamily()(pys-method),348park.ml.classification.LogisticRegressionsetK()(pyspark.ml.clustering.Gaus
sianMixture method),328 method),353
setFamily() (pys-setK()(pyspark.ml.clustering.KMeansmethod),
351
park.ml.regression.GeneralizedLinearRegression
setK()(pyspark.ml.clustering.PowerIterationClustering park.ml.classification.NaiveBayesmethod),

method),362
 setKeepLastCheckpoint() (pys-setNonnegative() 342
 park.ml.clustering.LDAMethod),357 park.ml.recommendation.ALSsetLayers() (pys- 366
 park.ml.classification.MultilayerPerceptronClassifiersetNumBlocks()
 method),345 park.ml.recommendation.ALS
 setLearningDecay() (pys- 366
 park.ml.clustering.LDAMethod),357 setNumFolds()
 setLearningOffset() (pys- park.ml.tuning.CrossValidatorpark.ml.clustering.LDAMethod),357 372
 setLink() (pys-setNumItemBlocks()
 park.ml.regression.GeneralizedLinearRegression park.ml.recommendation.ALSmethod),312 366
 setLinkPower() (pys-setNumUserBlocks()
 park.ml.regression.GeneralizedLinearRegression park.ml.recommendation.ALSmethod),312 366
 setLinkPredictionCol() (pys-setOffsetCol()
 (pysmethod),
 (pysmethod),
 (pysmethod),
 (pysmethod),
 (pysmethod),
 (pys-
 park.ml.regression.GeneralizedLinearRegression park.ml.regression.GeneralizedLinearRegression
 method),312 method),312
 setLossType() (pys-setOptimizeDocConcentration()(pyspark.ml.clustering.LDAMethod),357setOptimizer()(pyspark.ml.clustering.LDAMethod),357
 park.ml.classification.GBTClassifiermethod),337 setParams() (pys-
 setLossType() (pys- park.ml.classification.DecisionTreeClassifiermethod),335
 park.ml.regression.GBTRegressormethod), setParams() (pys-
 308setLowerBoundsOnCoefficients()(pys- park.ml.classification.GBTClassifiermethod),337
 park.ml.classification.LogisticRegressionmethod),328 setParams() (pys-
 setLowerBoundsOnIntercepts() (pys- park.ml.classification.LinearSVCmethod),
 park.ml.classification.LogisticRegressionmethod),328 325
 setMetricName() (pys-
 park.ml.evaluation.BinaryClassificationEvaluatorsetParams()(pysmethod),376park.ml.classification.LogisticRegressionsetMetricName()(pys-method),328
 park.ml.evaluation.ClusteringEvaluatorsetParams()(pysmethod),379park.ml.classification.MultilayerPerceptronClassifiersetMetricName()(pys-method),345
 park.ml.evaluation.MulticlassClassificationEvaluatorsetParams()(pysmethod),378park.ml.classification.NaiveBayesmethod),setMetricName()(pys-342
 park.ml.evaluation.RegressionEvaluatorsetParams()(pysmethod),377park.ml.classification.OneVsRestmethod),
 setMinDivisibleClusterSize()(pys-346park.ml.clustering.BisectingKMeanssetParams()(pysmethod),349park.ml.classification.RandomForestClassifier
 setModelType() (pys- method),340
 setParams() (pys- park.ml.regression.RandomForestRegressor
 park.ml.clustering.BisectingKMeansmethod),323method),349setParams()(pyspark.ml.tuning.CrossValidatorsetParams()(pys-method),372
 park.ml.clustering.GaussianMixturesetParams()(pysmethod),353park.ml.tuning.TrainValidationSplit
 setParams() (pyspark.ml.clustering.KMeans method),374
 method),351 setQuantileProbabilities() (pys-
 setParams() (pyspark.ml.clustering.LDA park.ml.regression.AFTSurvivalRegressionmethod),357 method),304
 setParams() (pys-setQuantilesCol() (pys-
 park.ml.clustering.PowerIterationClustering park.ml.regression.AFTSurvivalRegressionmethod),362 method),304
 setParams()(pys-setRank()(pyspark.ml.recommendation.ALSpark.ml.evaluation.BinaryClassificationEvaluatormethod),366
 method),377 setRatingCol() (pyssetParams() (pys- park.ml.recommendation.ALS method),
 park.ml.evaluation.ClusteringEvaluator 366
 method),379setSmoothing()(pyssetParams()(pys-park.ml.classification.NaiveBayesmethod),park.ml.evaluation.MulticlassClassificationEvaluator342

Index

attribute),318summary()
(pyspark.ml.stat.SummaryBuildermethod),302
SummaryBuilder(classinpyspark.ml.stat),302
T
test()(pyspark.ml.stat.ChiSquareTest staticmethod),297
test()(pyspark.ml.stat.KolmogorovSmirnovTeststaticmethod),299
theta(pyspark.ml.classification.NaiveBayesModelattribute),342
toLocal()
park.ml.clustering.DistributedLDAModelmethod),360
topicsMatrix()
359 park.ml.clustering.LDAModel
totalIterations (pys-

userFactors (pys-
park.ml.recommendation.ALSModelattribute),368
V
validateStages() (pys-
park.ml.pipeline.PipelineSharedReadWritestaticmethod),370
validationMetrics (pys-
park.ml.tuning.TrainValidationSplitModelattribute),375
(pys-variance()(pyspark.ml.stat.Summarizerstaticmethod),302vocabSize()(pyspark.ml.clustering.LDAModelmethod),35
9
(pys-W
method),weightedFalsePositiveRate
(pys-
park.ml.classification.LogisticRegressionSummary

park.ml.classification.LogisticRegressionTrainingSummaryattribute),331attribute),332
weightedFMeasure() (pys-
park.ml.classification.LogisticRegressionSummarymethod),331
weightedPrecision (pys-
park.ml.classification.LogisticRegressionSummaryattribute),331
weightedRecall (pys-
park.ml.classification.LogisticRegressionSummaryattribute),331
weightedTruePositiveRate (pys-
park.ml.classification.LogisticRegressionSummaryattribute),331
weights(pyspark.ml.classification.MultilayerPerceptronClassificationModelattribute),345
weights(pyspark.ml.clustering.GaussianMixtureModelattribute),354
write()(pyspark.ml.pipeline.Pipelinemethod),
369write() (pyspark.ml.pipeline.PipelineModelmethod),369
write() (pyspark.ml.tuning.CrossValidatormethod),372
write()(pyspark.ml.tuning.CrossValidatorModelmethod),373
write()(pyspark.ml.tuning.TrainValidationSplitmethod),374
write()(pyspark.ml.tuning.TrainValidationSplitModelmethod),375

Viewpublicationstats