Name	Formula	Direction	Range	Description
Performance measures for regression				
Mean Squared Error (MSE)	$\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\left(y^{(i)}-\hat{y}^{(i)}\right)^2$	min	$[0,\infty)$	Mean of the squared distances between the target variable y and the predicted target $\hat{y}$.
Mean Absolute Error (MAE)	$\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\left\lvert y^{(i)}-\hat{y}^{(i)}\right\rvert$	min	$[0,\infty)$	More robust than MSE, since it is less influenced by large errors.
R^2	$1-\frac{\sum_{i=1}^{n_{\text{test}}}\left(y^{(i)}-\hat{y}^{(i)}\right)^2}{\sum_{i=1}^{n_{\text{test}}}\left(y^{(i)}-\bar{y}\right)^2}$	max	$(-\infty,1]$	Compare the sum of squared errors (SSE) of the model to a constant baseline model.
Performance measures for classification based on class labels				
Accuracy (ACC)	$\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\mathbb{I}_{\left\{y^{(i)}=\hat{y}^{(i)}\right\}}$	max	$[0,1]$	Proportion of correctly classified observations.
Balanced Accuracy (BA)	$\frac{1}{g}\sum_{k=1}^{g}\frac{1}{n_{\text{test},k}}\sum_{y^{(i)}:y^{(i)}=k}\mathbb{I}_{\left\{y^{(i)}=\hat{y}^{(i)}\right\}}$	max	$[0,1]$	Variant of the accuracy that accounts for imbalanced classes.
Classification Error (CE)	$\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\mathbb{I}_{\left\{y^{(i)}\neq\hat{y}^{(i)}\right\}}$	min	$[0,1]$	CE $= 1 -$ ACC is the proportion of incorrect predictions.
ROC measures	$\text{TPR}=\frac{\text{TP}}{\text{TP+FN}}$	max	$[0,1]$	True Positive Rate: how many observations of the positive class 1 are predicted as 1?
	$\text{FPR}=\frac{\text{FP}}{\text{TN+FP}}$	min	$[0,1]$	False Positive Rate: how many observations of the negative class 0 are falsely predicted as 1?
	$\text{TNR}=\frac{\text{TN}}{\text{TN+FP}}$	max	$[0,1]$	True Negative Rate: how many observations of the negative class 0 are predicted as 0?
	$\text{FNR}=\frac{\text{FN}}{\text{TP+FN}}$	min	$[0,1]$	False Negative Rate: how many observations of the positive class 1 were falsely predicted as 0?
	$\text{PPV}=\frac{\text{TP}}{\text{TP+FP}}$	max	$[0,1]$	Positive Predictive Value: how likely is a predicted 1 a true 1?
	$\text{NPV}=\frac{\text{TN}}{\text{FN+TN}}$	max	$[0,1]$	Negative Predictive Value: how likely is a predicted 0 a true 0?
F_1	$2\frac{\text{PPV}\cdot\text{TPR}}{\text{PPV+TPR}}$	max	$[0,1]$	F_1 is the harmonic mean of PPV and TPR. Especially useful for imbalanced classes.
Cost measure	$\sum_{i=1}^{n_{\text{test}}}C(y^{(i)},\hat{y}^{(i)})$	min	$[0,\infty)$	Cost of incorrect predictions based on a (usually non-negative) cost matrix $C\in\mathbb{R}^{g,g}$.
Performance measures for classification based on class probabilities				
Brier Score (BS)	$\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\sum_{k=1}^{g}\left(\hat{\pi}_k(\mathbf{x}^{(i)})-\sigma_k(y^{(i)})\right)^2$	min	$[0,1]$	Measures squared distances of probabilities from the one-hot encoded class labels.
Log-Loss (LL)	$\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\left(-\sum_{k=1}^{g}\sigma_k(y^{(i)})\log(\hat{\pi}_k(\mathbf{x}^{(i)}))\right)$	min	$[0,\infty)$	A.k.a. Bernoulli, binomial or cross-entropy loss
AUC		max	$[0,1]$	Area under the ROC curve.

$\hat{y}^{(i)}$ denotes the predicted label for observation $\mathbf{x}^{(i)}$. ACC, BA, CE, BS, and LL can be used for multi-class classification with g classes. For AUC, multiclass extensions exist as well. The notation $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function. $\sigma_k(y)=\mathbb{I}_{\{y=k\}}$ is 1 if y is class k, 0 otherwise (multi-class one-hot encoding). $n_{\text{test},k}$ is the number of observations in the test set with class k. $\hat{\pi}_k(\mathbf{x})$ is the estimated probability for observation $\mathbf{x}^{(i)}$ of belonging to class k. TP is the number of true positives (observations of class 1 with predicted class 1), FP is the number of false positives (observations of class 0 with predicted class 1), TN is the number of true negatives (observations of class 0 with predicted class 0), and FN is the number of false negatives (observations of class 1 with predicted class 0).

Table 1: Popular performance measures used for ML, assuming an arbitrary test set of size n_{test}.