

## Lab 6 - Evaluation and Generalization

Francisco Freiria 97236 & João Morais 83916

Neste Laboratório foi proposto a fazer a avaliação de dois classificadores dados dois conjuntos de datasets, por forma a classificar os dois conjuntos de datasets, decidimos utilizar Naive Bayes Classifier e Support Vectors Machines como classificadores. Como medidas de performance utilizaram-se: accuracy e a precision.

Por forma a calcular as medidas de performance atrás referidas, recorreu-se ao conceito de confusion matrix. A confusion matrix produz o seguinte comportamento:

	<i>Actual</i>		
		<b>Positive - 1</b>	<b>Negative - 0</b>
	<b>Positive - 1</b>	True Positive	False Positive
	<b>Negative - 0</b>	False Negative	True Negative

- **True Positive** – Classe real do ponto atual é 1 e o classificador atribuí classe 1 também.

- **True Negative** – Classe real do ponto atual é 0 e o classificador atribuí classe 0 também.

- **False Positive** – Classe real do ponto atual é 0 e o classificador atribuí classe 1.

- **False Negative** – Classe real do ponto atual é 1 e o classificador atribuí classe 0.

As medidas de performance utilizadas foram:

$$\text{- Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

$$\text{- Precision} = \frac{TP}{TP+FP}$$

$$\text{- Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{- Specificity} = \frac{TN}{FP+TN}$$

Realizamos um programa onde testamos os dois datasets dados para os dois classificadores atrás referidos. Após este processo, são calculadas as medidas de performance atrás expostas. Na tabela seguinte são apresentados os resultados obtidos:

DATASET	1		2	
Classificador	Naives Bayes	Suport Vector Machine (SVM)	Naives Bayes	Suport Vector Machine (SVM)
Accuracy	0.513	0.765	0.758	0.838
Precision	0.481	0.870	0.774	0.774
Sensitivity	0.481	0.7014	0.585	0.727
Specificity	0.541	0.854	0.879	0.894

Analisando os resultados expostos na tabela 2, conclui-se que:

- No SVM foi usada a função polinomial e foram conseguidos bons resultados. Foram também testados vários graus do polinómio de Kernel como medida de otimização. O melhor valor conseguido para o grau do polinómio foi o polinómio de primeiro grau para o dataset1 e o polinómio de grau 4 para o dataset2. Foi tomado também em consideração a avaliação do score e consequentemente do erro.

- No que toca ao parâmetro precision, os classificadores Naive Bayes e SVM apresentam valores semelhantes no dataset2;

- O SVM apresenta-se como o classificador com mais sensitivity;

- O classificador SVM é o que apresenta maior specificity.

- Em termos de accuracy, quando comparando os dois datasets, concluímos que o classificador SVM é o que apresenta melhores resultados;

- Para o dataset1 podemos que o Naives Bayes não é um bom classificador apresentando apenas uma accuracy de 51%.

Face ao acima exposto, é fácil de concluir que os diferentes classificadores apresentam diferentes características. Ainda assim, o classificador Support Vectors Machines(SVM) apresenta-se como o classificador mais consistente dos estudados.