

Resumen

Este documento resume brevemente los resultados obtenidos de la implementación de un clasificador de los números entre 0 y 9 a partir de imágenes.

Autores: Andrés Gómez, Sebastián Ruiz

Parte I

Descripción de la solución

El clasificador de números está compuesto por 10 clasificadores binarios en una configuración uno contra uno; de forma que cada uno de los clasificadores se entrena para reconocer un dígito en específico del resto. Los clasificadores son modelados con base en la regresión lineal; así que, se puede obtener la probabilidad de una determinada muestra de ser el número con el que el clasificador se ha entrenado. La etiqueta del número final es entonces, aquella que tenga la probabilidad más alta entre todos los clasificadores.

Parte II

Resultados de la solución

1. Clasificadores individuales

El entrenamiento del clasificador de números radica en los diez clasificadores logísticos que la componen. Para el entrenamiento de cada clasificador se emplean como hiperparámetros: la tasa de aprendizaje, el número de iteraciones y el coeficiente de regularización. Como métricas de aprendizaje se emplearon la exactitud y el puntaje f1. La meta del entrenamiento fue conseguir que el desempeño sobre el conjunto de entrenamiento fuera similar al desempeño sobre conjunto de prueba para evitar el sobreentrenamiento. A continuación se muestran los resultados tanto del conjunto de entrenamiento Cuadro 1 como el conjunto de prueba Cuadro 2.

	Métrica exactitud	Métrica F1
Clasificador 0	0.992	0.978
Clasificador 1	0.993	0.984
Clasificador 2	0.981	0.946
Clasificador 3	0.975	0.931
Clasificador 4	0.981	0.957
Clasificador 5	0.977	0.926
Clasificador 6	0.987	0.965
Clasificador 7	0.986	0.963
Clasificador 8	0.950	0.863
Clasificador 9	0.963	0.894

Cuadro 1: Clasificadores de entrenamiento

	Métrica exactitud	Métrica F1
Clasificador 0	0.993	0.981
Clasificador 1	0.993	0.983
Clasificador 2	0.981	0.946
Clasificador 3	0.976	0.931
Clasificador 4	0.983	0.953
Clasificador 5	0.978	0.928
Clasificador 6	0.986	0.959
Clasificador 7	0.984	0.956
Clasificador 8	0.947	0.853
Clasificador 9	0.963	0.895

Cuadro 2: Clasificadores de prueba

2. Clasificador global

El desempeño global del sistema puede resumirse en la matriz de confusión, la exactitud y el puntaje f1 promedio en el conjunto de entrenamiento y prueba. En la siguiente tabla Cuadro 3 se muestran los resultados de la exactitud y el puntaje F1 de todo el sistema.

	Métrica exactitud	Métrica F1
Clasificadores prueba	0.918	0.917
Clasificadores entrenamiento	0.922	0.921

Cuadro 3: Clasificadores comparación global

Parte III

Conclusiones

Los resultados del clasificador se consideran que tuvieron un desempeño bueno. Dada la comparación hecha, en los clasificadores individuales y el clasificador global, se puede observar que hubo cambios mínimos; lo que consideramos como muy bajo sobreentrenamiento. Además, aunque el tiempo del entrenamiento fue demorado, se encontró al final buenos porcentajes tanto para el conjunto de entrenamiento como para el conjunto de prueba.

Este sistema es una forma ingeniosa de emplear la habilidad de la regresión logística de calcular la probabilidad de pertenencia a una clase. Esto nos otorga cierta trazabilidad y comprensión acerca de cómo funciona el modelo.

Uno de los factores que pueden estar afectando el desempeño del sistema es el desbalance que existe entre el número de muestras de un solo número contra el de las muestras de todas las etiquetas. Otro factor importante es la fragilidad del sistema: dado que existen múltiples clasificadores, el pobre entrenamiento de tan solo uno de los clasificadores puede tener importante impacto sobre el desempeño final.