

## ENTRENAR UN MODELO

### Características de la Escritura Manual

El archivo MAT `datosletras.mat` contiene la tabla `datos_entrena`, que representa los datos de las características de 2906 muestras de letras individuales. Hay 25 características, incluidas medidas estadísticas, correlaciones y máximos/mínimos para la posición, velocidad y presión del lápiz.

Features

	1	2	3	4	5	6	7	8	9	10	11		16	17	18	19	20	21	22	23	24	25	26
	AspectRatio	MADx	MADy	Angle	MADU	AngleU	MADV	ConcV	ConcP	ConcU	ConcD		varPV	ConcV	NumCMin	NumCMax	NumCMin	NumCMax	NumCMin	NumCMax	NumCMin	NumCMax	Character
1	0.4816	0.4516	0.4991	-0.3279	-0.4547	-0.4547	0.4547	0.4547	0.4547	0.4547	0.4547		0.4547	0.4547	0.4547	0.4547	0.4547	0.4547	0.4547	0.4547	0.4547	0.4547	A
2	0.6790	0.1515	0.1988	0.1155	0.0967	-1.7575	0.1885	-0.4151	0.4676	0.1021	0.0380		-0.1858	0.3672	0	1	1	2	1	1	1	2	B
3	1.8667	0.2237	0.1512	0.0772	1.7257	-1.3183	1.6812	0.1900	-0.8001	0.5962	0.9452		-0.6721	0.5962	1	0	0	1	1	1	1	1	C
4	2.9476	0.1838	0.3128	0.2030	1.2513	-0.5708	0.5985	0.5130	0.0219	0.2745	0.2128		0.0602	0.5188	1	0	1	2	2	1	1	2	D
5	4.4245	0.1485	0.0540	0.3152	0.8110	-2.2975	0.1980	-0.5570	0.0253	0.0899	0.1211		0.3875	0.0586	0	0	1	1	0	1	1	2	E
6	0.9574	0.0874	0.7344	0.2623	0.0304	-1.4520	0.1755	-0.2079	-0.4036	0.7105	0.4687		-0.7944	-0.0627	0	1	1	0	2	1	1	1	F
7	0.3335	0.1180	0.5151	0.2762	1.2237	-1.7750	0.3113	0.1448	-0.3252	0.2530	0.3838		0.0236	0.0308	2	1	1	1	2	1	2	1	G
8	5.0000	0.1436	0.6333	0.3839	0.7443	-0.5740	4.2003	0.8799	0.6117	0.2113	0.2426		0.2235	0.5109	1	1	0	1	1	1	1	1	H
9	1.6852	0.2706	0.3666	0.7544	0.7140	-1.8837	0.5797	0.1291	0.1376	0.4763	0.1680		0.6526	0.6171	0	0	1	1	1	1	2	1	I
10	1.0000	0.2339	0.3021	0.0551	2.1183	0.0661	2.1855	0.2212	-0.0480	0.3980	0.9544		0.0601	0.2363	1	1	1	1	2	1	1	1	J
11	2.6875	0.1019	0.4334	0.1801	1.3070	-1.2440	2.7469	-0.0873	-0.3546	0.2863	0.4724		-0.3824	0.5404	2	2	2	2	2	2	2	2	K
12	1.8885	0.2136	0.3388	0.7543	0.0883	0.9304	2.2205	0.7719	-0.2512	0.8487	0.5882		-0.0285	0.0320	0	0	1	0	0	1	1	1	L
13	1977	0.5444	-0.0265	1.2599	-1.5230	1.2687	1.2687	0.4082	0.0625	0.0545	-0.1862		-0.7512	-0.5458	1	0	0	0	2	1	1	1	M
14	2533	0.3842	0.6607	0.8040	-1.2508	3.7354	-0.2025	0.2430	0.4095	0.3634			0.1620	0.2946	0	0	1	1	2	1	2	1	N
15	5481	0.4205	0.8334	0.4186	-0.0061	0.7186	-0.2386	0.3475	0.5762	0.7174			-0.1470	0.2862	0	0	1	0	0	1	1	1	O
16	3362	0.3116	1.0408	0.6083	0.1361	2.6785	0.6426	0.1346	0.4249	0.6187			-0.0752	0.3788	0	0	2	1	1	1	2	2	P
17	3408	0.4706	0.6940	2.8877	-4.4812	3.7626	-0.1222	0.0716	-0.1496	0.5774			-0.4280	-0.4178	0	1	1	1	1	1	1	2	Q
18	3076	0.4771	0.1183	1.1258	-2.0738	2.9101	-0.0682	0.6845	0.0640	0.0907			-0.1887	0.4821	0	2	1	1	2	2	2	2	R
19	1612	0.2953	0.4579	1.6404	-0.7401	3.5321	-0.1161	0.2826	0.3639	0.4891			0.0413	0.0871	1	0	2	2	6	6	6	2	S
20	1746	0.3377	0.1303	1.3938	-1.8105	2.3869	-0.3205	0.6010	-0.0308	-0.0201			0.2771	0.1680	0	1	1	1	3	3	3	5	T
21	2236	0.4528	0.7859	1.1504	-1.6895	5.6487	-0.2218	-0.1286	0.3936	0.6544			0.4182	0.3870	1	0	1	1	2	2	2	4	U
22	3453	0.2881	0.7946	2.5339	-0.5391	2.1984	0.2191	0.1709	0.1764	0.8841			0.1284	0.2385	1	1	0	1	1	2	3	1	V
23	1315	0.3719	1.4552	1.3838	0.3558	0.8838	0.4993	0.0381	0.2877	-0.6400			0.2886	-0.7253	2	1	1	1	6	6	6	7	W
24	1519	0.2627	0.6273	1.6770	-0.7562	3.3238	-0.1910	0.3439	0.2688	0.4415			0.5180	0.5072	2	1	2	2	1	1	2	3	X
25	1916	0.4127	0.7905	0.6842	-1.7779	3.2288	-0.3719	0.3207	0.2257	-0.0238			-0.0915	-0.8457	0	0	1	1	1	1	2	3	Y
26	3589	1.3964	0.3935	0.4248	-0.0515	5.7018	-0.3823	0.4410	0.0244	0.1775			-0.4805	-0.1961	0	1	1	2	3	3	3	3	Z
27	2.5945	0.1821	0.3039	-0.5626	0.0752	-1.4145	2.1030	0.3542	-0.0883	0.4385	-0.9146		0.4230	-0.6809	0	1	1	0	2	2	2	1	aa
28	2.4874	0.2382	0.4400	1.2273	1.1264	-4.5044	0.9081	0.0775	0.2320	0.7963	0.8444		0.0278	0.7519	1	1	1	1	2	3	2	2	bb
29	45.6887	0.0427	2.3918	-0.0776	0.3714	-7.1032	5.6575	0.5875	-0.3844	0.0775	-0.3803		-0.3807	0.0231	0.0000	1	1	1	2	2	2	2	cc
30	2.2833	0.2623	0.4573	0.7541	0.9132	-1.6987	0.9179	-0.4938	0.5327	0.1863	-0.4738		-0.0182	0.0002	0	0	1	1	1	2	2	2	dd
31	1.2833	0.1885	0.3107	-0.1018	1.0580	0.2020	2.5619	0.0777	0.3051	0.1885	0.9425		0.3037	0.1130	1	1	1	1	2	1	1	1	ee
32	3.7333	0.1290	0.4815	0.0234	1.1596	-0.8379	4.8819	0.4381	0.2970	0.0487	-0.3315		0.2237	0.2601	1	1	1	1	4	4	4	2	ff
33	1.6762	0.2764	0.3947	0.7405	0.8834	0.4515	2.4883	0.3884	0.3968	-0.2838	0.8742		0.4715	-0.3847	0	1	1	0	1	1	1	1	gg
34	2.4516	0.1855	0.3056	0.7689	0.0115	0.4796	5.5152	0.1100	0.5711	-0.3079	0.6878		0.4680	-0.2937	0	0	1	0	2	2	2	2	hh
35	1.2882	0.2431	0.3148	0.0334	0.3257	-0.1946	4.0620	0.0815	0.5129	0.1881	0.9838		0.3570	-0.1540	0	0	2	1	1	2	3	2	ii
36	1.4814	0.1977	0.2328	1.8889	0.4858	-2.8135	4.8496	-0.1401	0.2882	0.4887	0.0821		-0.2680	0.6327	2	3	1	2	3	3	3	2	jj
37	1.9886	0.2474	0.3245	0.7223	1.2337	-1.2647	0.8462	-0.3390	0.4215	0.4462	0.4613		-0.3887	0.3289	1	0	2	2	2	2	2	2	kk
38	1.2252	0.1881	0.3477	0.5196	1.3827	-1.0882	3.2142	-0.2101	-0.0705	0.5583	0.3930		-0.2817	0.3296	1	0	2	2	2	2	2	2	ll
39	2.0714	0.1885	0.3821	0.2257	1.3488	-1.4297	3.0829	-0.4049	0.0091	0.0412	-0.1564		0.3483	0.0689	0	1	2	2	1	2	2	2	mm
40	7.4887	0.1706	0.4814	0.1931	1.0467	-3.7519	1.4165	-0.4844	0.4363	0.0454	0.1546		0.3463	0.0844	0	1	3	1	1	1	1	3	nn

### Actividad 1

Los datos de entrenamiento pueden introducir sesgos en su modelo. ¿Cómo se distribuyen las 26 letras del alfabeto en los datos de entrenamiento de escritura manual?

**Tarea:** Cargue los datos:

```
load datosletras.mat
```

Utilice la función `histogram` para crear un histograma de las clases de letras conocidas (almacenadas en la variable `Character` de la tabla `datos_entrena`).

### Actividad 2

Un diagrama de caja es una forma sencilla de visualizar varias distribuciones.

```
boxplot(x,c)
```

Esto crea un diagrama donde las cajas representan la distribución de los valores de `x` para cada una de las clases de `c`. Si los valores de `x` suelen ser significativamente diferentes para una clase en comparación con otra, entonces `x` es una característica que puede distinguir entre esas clases. Cuantas más características tenga que puedan distinguir las diferentes clases, más posibilidades tendrá de crear un modelo de clasificación preciso a partir del conjunto de datos completo.

**Tarea:** Utilice la función `boxplot` para crear un diagrama de caja de los valores de la característica `MADX` (desviación media absoluta de la posición horizontal) para cada letra. Las clases de letras conocidas se almacenan en la variable llamada `Character`.

#### Tarea adicional

Use el comando `classificationLearner` para abrir la app Classification Learner.

- Seleccione `datos_entrena` como los datos que utilizar.
- La app debería detectar correctamente `Character` como la variable de respuesta que predecir.
- Elija la opción de validación predeterminada.
- Seleccione un modelo y haga clic en el botón **Train**.

Pruebe algunos de los modelos estándar con las opciones predeterminadas. **Observe si puede conseguir al menos un 80% de precisión.**

Tenga en cuenta que las SVM funcionan en problemas de clasificación binaria (es decir, cuando solo hay dos clases). Para hacer que las SVM funcionen en este problema, la app se ajusta a muchas SVM. Por lo tanto, estos modelos serán lentos de entrenar.

#### Archivos requeridos:

`datosletras.mat`