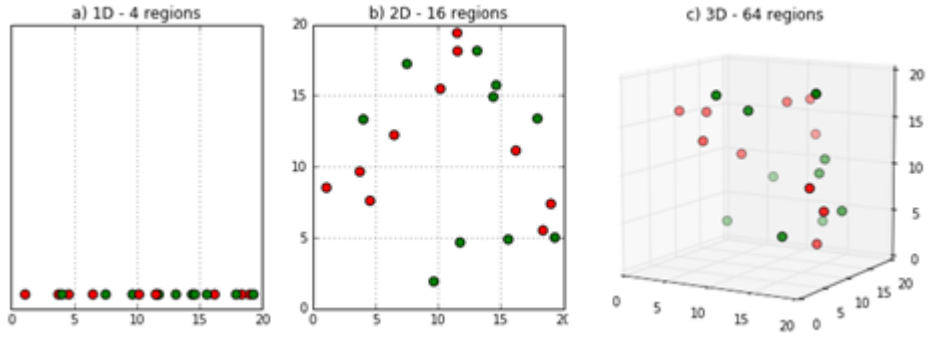

THE BRIDGE
No Supervisado
Reducción de Dimensionalidad
(Extensión)
PCA

Curse of Dimensionality

Curse of Dimensionality

El número de muestras que se necesitan para estimar una función arbitraria (un target de ML, por ejemplo) con un cierto nivel de precisión crece exponencialmente con el número de inputs/dimensiones/variables de la función.

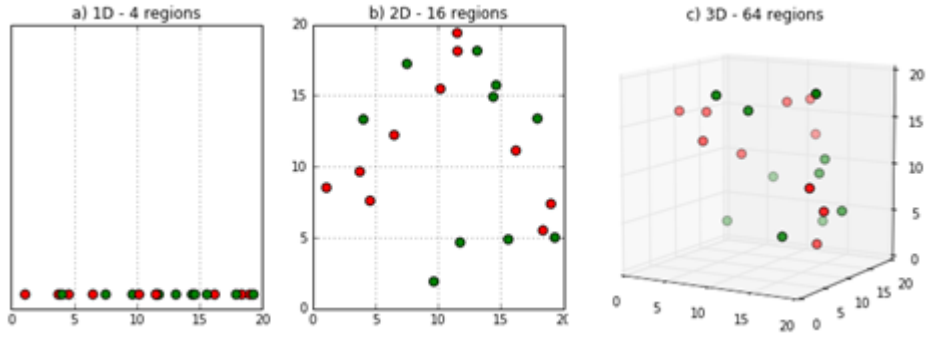
Este fenómeno afecta mucho a la dispersión y la cercanía de los datos. Según vamos añadiendo dimensiones, se van diferenciando mejor.



Curse of Dimensionality

El número de muestras que se necesitan para estimar una función arbitraria (un target de ML, por ejemplo) con un cierto nivel de precisión crece exponencialmente con el número de inputs/dimensiones/variables de la función.

Este fenómeno afecta mucho a la dispersión y la cercanía de los datos. Según vamos añadiendo dimensiones, se van diferenciando mejor.

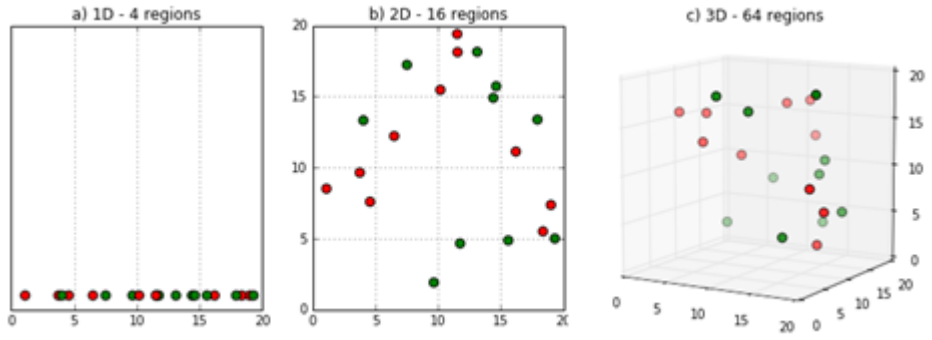


Cuando estamos ante pocas dimensiones, tenemos datos que pueden resultar muy parecidos, pero según vamos añadiendo características y dimensiones nuevas, esto cambia

Curse of Dimensionality

El número de muestras que se necesitan para estimar una función arbitraria (un target de ML, por ejemplo) con un cierto nivel de precisión crece exponencialmente con el número de inputs/dimensiones/variables de la función.

Este fenómeno afecta mucho a la dispersión y la cercanía de los datos. Según vamos añadiendo dimensiones, se van diferenciando mejor.



Cuando estamos ante pocas dimensiones, tenemos datos que pueden resultar muy parecidos, pero según vamos añadiendo características y dimensiones nuevas, esto cambia

Datasets con muchas dimensiones serán muy dispersos y con mucha distancia entre los puntos, lo cual es bueno para clasificar. El problema es que nuevas observaciones estarán también muy lejanas de las originales (**overfitting**), produciendo predicciones menos fiables que datasets con pocas dimensiones. **La solución sería incrementar el conjunto de train**.

Reducir la dimensionalidad: Reducir las “features”

Reducir la dimensionalidad: Reducir las "features"

Datos Originales
(pej: 4096 features)

Valor informativo = M

	pixel_0	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	pixel_7	pixel_8	pixel_9	...	pixel_4086	pixel_4087	pixel_4088	pixel_4089	pixel_4090	pixel_4091	pixel_4092	pixel_4093
0	0.309917	0.367769	0.417355	0.442149	0.528926	0.607438	0.657025	0.677686	0.690083	0.685950	...	0.665289	0.669421	0.652893	0.661157	0.475207	0.132231	0.148760	0.152893
1	0.454545	0.471074	0.512397	0.557851	0.595041	0.640496	0.681818	0.702479	0.710744	0.702479	...	0.136364	0.157025	0.136364	0.148760	0.152893	0.152893	0.152893	0.152893
2	0.318182	0.400826	0.491736	0.528926	0.586777	0.657025	0.681818	0.685950	0.702479	0.698347	...	0.074380	0.132231	0.181818	0.136364	0.128099	0.148760	0.144628	0.140496
3	0.198347	0.194215	0.194215	0.194215	0.190083	0.190083	0.243802	0.404959	0.483471	0.516529	...	0.652893	0.636364	0.657025	0.685950	0.727273	0.743802	0.764463	0.752066
4	0.500000	0.545455	0.582645	0.623967	0.648760	0.690083	0.694215	0.714876	0.723140	0.731405	...	0.190083	0.161157	0.177686	0.173554	0.177686	0.177686	0.177686	0.177686
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
395	0.400826	0.495868	0.570248	0.632231	0.648760	0.640496	0.661157	0.636364	0.665289	0.698347	...	0.388430	0.396694	0.264463	0.099174	0.181818	0.243802	0.247934	0.161157
396	0.367769	0.367769	0.351240	0.301653	0.247934	0.247934	0.367769	0.512397	0.574380	0.628099	...	0.380165	0.334711	0.289256	0.285124	0.338843	0.404959	0.458678	0.487603
397	0.500000	0.533058	0.607438	0.628099	0.657025	0.632231	0.657025	0.669421	0.673554	0.702479	...	0.194215	0.148760	0.152893	0.161157	0.161157	0.173554	0.157025	0.177686
398	0.214876	0.219008	0.219008	0.223140	0.210744	0.202479	0.276859	0.400826	0.487603	0.549587	...	0.446281	0.392562	0.367769	0.409091	0.479339	0.524793	0.545455	0.574380
399	0.516529	0.462810	0.280992	0.252066	0.247934	0.367769	0.574380	0.615702	0.661157	0.615702	...	0.276859	0.264463	0.293388	0.301653	0.293388	0.322314	0.322314	0.359504

Reductor de Dimensionalidad (by Fluzo Corp.)

Reducir la dimensionalidad: Reducir las "features"

Datos Originales
(pej: 4096 features)

Valor informativo = M

	pixel_0	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	pixel_7	pixel_8	pixel_9	...	pixel_4086	pixel_4087	pixel_4088	pixel_4089	pixel_4090	pixel_4091	pixel_4092	pixel_4093
0	0.309917	0.367769	0.417355	0.442149	0.528926	0.607438	0.657025	0.677686	0.690083	0.685950	...	0.665289	0.669421	0.652893	0.661157	0.475207	0.132231	0.148760	0.152893
1	0.454545	0.471074	0.512397	0.557851	0.595041	0.640496	0.681818	0.702479	0.710744	0.702479	...	0.136364	0.157025	0.136364	0.148760	0.152893	0.152893	0.152893	0.152893
2	0.318182	0.400826	0.491736	0.528926	0.586777	0.657025	0.681818	0.685950	0.702479	0.698347	...	0.074380	0.132231	0.181818	0.136364	0.128099	0.148760	0.144628	0.140496
3	0.198347	0.194215	0.194215	0.194215	0.190083	0.190083	0.243802	0.404959	0.483471	0.516529	...	0.652893	0.636364	0.657025	0.685950	0.727273	0.743802	0.764463	0.752066
4	0.500000	0.545455	0.582645	0.623967	0.648760	0.690083	0.694215	0.714876	0.723140	0.731405	...	0.190083	0.161157	0.177686	0.173554	0.177686	0.177686	0.177686	0.177686
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
395	0.400826	0.495868	0.570248	0.632231	0.648760	0.640496	0.661157	0.636364	0.665289	0.698347	...	0.388430	0.396694	0.264463	0.099174	0.181818	0.243802	0.247934	0.161157
396	0.367769	0.367769	0.351240	0.301653	0.247934	0.247934	0.367769	0.512397	0.574380	0.628099	...	0.380165	0.334711	0.289256	0.285124	0.338843	0.404959	0.458678	0.487603
397	0.500000	0.533058	0.607438	0.628099	0.657025	0.632231	0.657025	0.669421	0.673554	0.702479	...	0.194215	0.148760	0.152893	0.161157	0.161157	0.173554	0.157025	0.177686
398	0.214876	0.219008	0.219008	0.223140	0.210744	0.202479	0.276859	0.400826	0.487603	0.549587	...	0.446281	0.392562	0.367769	0.409091	0.479339	0.524793	0.545455	0.574380
399	0.516529	0.462810	0.280992	0.252066	0.247934	0.367769	0.574380	0.615702	0.661157	0.615702	...	0.276859	0.264463	0.293388	0.301653	0.293388	0.322314	0.322314	0.359504

Reductor de Dimensionalidad (by Fluzo Corp.)

Reducir la dimensionalidad: Reducir las "features"

Datos Originales
(pej: 4096 features)

Valor informativo = M

	pixel_0	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	pixel_7	pixel_8	pixel_9	...	pixel_4086	pixel_4087	pixel_4088	pixel_4089	pixel_4090	pixel_4091	pixel_4092	pixel_4093
0	0.309917	0.367769	0.417355	0.442149	0.528926	0.607438	0.657025	0.677686	0.690083	0.685950	...	0.665289	0.669421	0.652893	0.661157	0.475207	0.132231	0.148760	0.152893
1	0.454545	0.471074	0.512397	0.557851	0.595041	0.640496	0.681818	0.702479	0.710744	0.702479	...	0.136364	0.157025	0.136364	0.148760	0.152893	0.152893	0.152893	0.152893
2	0.318182	0.400826	0.491736	0.528926	0.586777	0.657025	0.681818	0.685950	0.702479	0.698347	...	0.074380	0.132231	0.181818	0.136364	0.128099	0.148760	0.144628	0.140496
3	0.198347	0.194215	0.194215	0.194215	0.190083	0.190083	0.243802	0.404959	0.483471	0.516529	...	0.652893	0.636364	0.657025	0.685950	0.727273	0.743802	0.764463	0.752066
4	0.500000	0.545455	0.582645	0.623967	0.648760	0.690083	0.694215	0.714876	0.723140	0.731405	...	0.190083	0.161157	0.177686	0.173554	0.177686	0.177686	0.177686	0.177686
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	0.400826	0.495868	0.570248	0.632231	0.648760	0.640496	0.661157	0.636364	0.665289	0.698347	...	0.388430	0.396694	0.264463	0.099174	0.181818	0.243802	0.247934	0.161157
996	0.367769	0.367769	0.351240	0.301653	0.247934	0.247934	0.367769	0.512397	0.574380	0.628099	...	0.380165	0.334711	0.289256	0.285124	0.338843	0.404959	0.458678	0.487603
997	0.500000	0.533058	0.607438	0.628099	0.657025	0.632231	0.657025	0.669421	0.673554	0.702479	...	0.194215	0.148760	0.152893	0.161157	0.161157	0.173554	0.157025	0.177686
998	0.214876	0.219008	0.219008	0.223140	0.210744	0.202479	0.276859	0.400826	0.487603	0.549587	...	0.446281	0.392562	0.367769	0.409091	0.479339	0.524793	0.545455	0.574380
999	0.516529	0.462810	0.280992	0.252066	0.247934	0.367769	0.574380	0.615702	0.661157	0.615702	...	0.276859	0.264463	0.293388	0.301653	0.293388	0.322314	0.322314	0.359504

Reductor de Dimensionalidad (by Fluzo Corp.)

Datos
Transformados
(pej: 200 features)

Valor informativo = M'

	feat_new_1	feat_new_2	feat_new_3	feat_new_4	feat_new_5	feat_new_6	feat_new_7	feat_new_8	feat_new_9	feat_new_10	...	feat_new_190	feat_new_191	feat_new_192	feat_new_193	feat_new_194
0	-4.083442	5.739435	-0.236276	0.578894	1.209647	0.082671	-1.802376	0.556551	-1.268074	-0.063922	...	0.009266	-0.199691	0.132035	-0.018869	0.089872
1	-2.046529	-6.809134	0.484901	-1.709170	3.854363	-0.926039	2.851492	-1.285769	1.284741	-2.956928	...	0.047114	-0.177876	0.071906	0.097713	0.307296
2	0.820682	0.851959	3.753963	3.759623	-1.550162	-1.923993	-0.192531	-1.136280	0.773947	0.185524	...	0.084868	0.026628	0.071523	-0.089659	0.052076
3	-4.174684	-0.149040	-1.316980	-1.183464	1.203103	-0.399595	0.895599	0.810825	1.386703	-0.086677	...	-0.024617	-0.145642	-0.015023	0.045428	0.206534
4	-2.121450	-1.247912	-2.280725	2.833535	-2.498348	-1.494220	1.627887	1.100612	-0.951477	-0.419119	...	-0.156896	0.118289	-0.160467	-0.072183	-0.279902
5	2.441739	-3.185948	0.549997	-0.163092	0.524658	0.327163	-1.025523	-0.644161	-2.552042	-1.288205	...	-0.159354	-0.053866	-0.334462	-0.239899	-0.083924
6	1.091135	-4.100239	2.306275	1.008786	0.691498	0.783061	-1.179359	-0.536694	1.745058	-0.140650	...	0.033374	0.094718	-0.062994	-0.062602	0.095937
7	-2.394622	3.678058	3.077595	-0.173826	-0.027975	-3.097763	-0.107871	-1.525856	0.583504	-1.393703	...	-0.148875	0.273172	0.145177	-0.067410	-0.184296
8	-2.696331	-4.404981	-3.000558	2.454991	0.173063	-0.731339	0.347564	-0.106078	1.566675	-0.234409	...	-0.129933	0.017554	0.001765	0.077345	-0.062335
9	5.177533	3.655297	-0.112525	0.310772	-1.065532	0.466331	0.652401	0.506439	-0.305994	-0.482194	...	0.009959	-0.039743	-0.082019	0.112407	0.116086
10	-5.168567	3.737079	3.309348	0.787426	1.185890	2.219284	0.844754	1.171803	1.409934	0.678537	...	0.260479	0.079164	-0.107387	0.008247	-0.114654
11	-1.153720	2.304289	1.353355	-0.659652	-0.219001	3.023613	-1.022502	0.531080	1.322670	-1.313569	...	-0.000985	0.169200	0.009480	0.042539	-0.069131

Reducir la dimensionalidad: Reducir las "features"

Datos Originales
(pej: 4096 features)

Valor informativo = M

	pixel_0	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	pixel_7	pixel_8	pixel_9	...	pixel_4086	pixel_4087	pixel_4088	pixel_4089	pixel_4090	pixel_4091	pixel_4092	pixel_4093
0	0.309917	0.367769	0.417355	0.442149	0.528926	0.607438	0.657025	0.677686	0.690083	0.685950	...	0.665289	0.669421	0.652893	0.661157	0.475207	0.132231	0.148760	0.152893
1	0.454545	0.471074	0.512397	0.557851	0.595041	0.640496	0.681818	0.702479	0.710744	0.702479	...	0.136364	0.157025	0.136364	0.148760	0.152893	0.152893	0.152893	0.152893
2	0.318182	0.400826	0.491736	0.528926	0.586777	0.657025	0.681818	0.685950	0.702479	0.698347	...	0.074380	0.132231	0.181818	0.136364	0.128099	0.148760	0.144628	0.140496
3	0.198347	0.194215	0.194215	0.194215	0.190083	0.190083	0.243802	0.404959	0.483471	0.516529	...	0.652893	0.636364	0.657025	0.685950	0.727273	0.743802	0.764463	0.752066
4	0.500000	0.545455	0.582645	0.623967	0.648760	0.690083	0.694215	0.714876	0.723140	0.731405	...	0.190083	0.161157	0.177686	0.173554	0.177686	0.177686	0.177686	0.177686
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	0.400826	0.495868	0.570248	0.632231	0.648760	0.640496	0.661157	0.636364	0.665289	0.698347	...	0.388430	0.396694	0.264463	0.099174	0.181818	0.243802	0.247934	0.161157
996	0.367769	0.367769	0.351240	0.301653	0.247934	0.247934	0.367769	0.512397	0.574380	0.628099	...	0.380165	0.334711	0.289256	0.285124	0.338843	0.404959	0.458678	0.487603
997	0.500000	0.533058	0.607438	0.628099	0.657025	0.632231	0.657025	0.669421	0.673554	0.702479	...	0.194215	0.148760	0.152893	0.161157	0.161157	0.173554	0.157025	0.177686
998	0.214876	0.219008	0.219008	0.223140	0.210744	0.202479	0.276859	0.400826	0.487603	0.549587	...	0.446281	0.392562	0.367769	0.409091	0.479339	0.524793	0.545455	0.574380
999	0.516529	0.462810	0.280992	0.252066	0.247934	0.367769	0.574380	0.615702	0.661157	0.615702	...	0.276859	0.264463	0.293388	0.301653	0.293388	0.322314	0.322314	0.359504

Reductor de Dimensionalidad (by Fluzo Corp.)

- Reducir complejidad: Ganar generalización
- Hacer más robusto
- Reducir tiempos
- "Mejorar" los modelos

Datos
Transformados
(pej: 200 features)

Valor informativo = M'

	feat_new_1	feat_new_2	feat_new_3	feat_new_4	feat_new_5	feat_new_6	feat_new_7	feat_new_8	feat_new_9	feat_new_10	...	feat_new_190	feat_new_191	feat_new_192	feat_new_193	feat_new_194
0	-4.083442	5.739435	-0.236276	0.578894	1.209647	0.082671	-1.802376	0.556551	-1.268074	-0.063922	...	0.009266	-0.199691	0.132035	-0.018869	0.089872
1	-2.046529	-6.809134	0.484901	-1.709170	3.854363	-0.926039	2.851492	-1.285769	1.284741	-2.956928	...	0.047114	-0.177876	0.071906	0.097713	0.307296
2	0.820682	0.851959	3.753963	3.759623	-1.550162	-1.923993	-0.192531	-1.136280	0.773947	0.185524	...	0.084868	0.026628	0.071523	-0.089659	0.052076
3	-4.174684	-0.149040	-1.316980	-1.183464	1.203103	-0.399595	0.895599	0.810825	1.386703	-0.086677	...	-0.024617	-0.145642	-0.015023	0.045428	0.206534
4	-2.121450	-1.247912	-2.280725	2.833535	-2.498348	-1.494220	1.627887	1.100612	-0.951477	-0.419119	...	-0.156896	0.118289	-0.160467	-0.072183	-0.279902
5	2.441739	-3.185948	0.549997	-0.163092	0.524658	0.327163	-1.025523	-0.644161	-2.552042	-1.288205	...	-0.159354	-0.053866	-0.334462	-0.239899	-0.083924
6	1.091135	-4.100239	2.306275	1.008786	0.691498	0.783061	-1.179359	-0.536694	1.745058	-0.140650	...	0.033374	0.094718	-0.062994	-0.062602	0.095937
7	-2.394622	3.678058	3.077595	-0.173826	-0.027975	-3.097763	-0.107871	-1.525856	0.583504	-1.393703	...	-0.148875	0.273172	0.145177	-0.067410	-0.184296
8	-2.696331	-4.404981	-3.000558	2.454991	0.173063	-0.731339	0.347564	-0.106078	1.566675	-0.234409	...	-0.129933	0.017554	0.001765	0.077345	-0.062335
9	5.177533	3.655297	-0.112525	0.310772	-1.065532	0.466331	0.652401	0.506439	-0.305994	-0.482194	...	0.009959	-0.039743	-0.082019	0.112407	0.116086
10	-5.168567	3.737079	3.309348	0.787426	1.185890	2.219284	0.844754	1.171803	1.409934	0.678537	...	0.260479	0.079164	-0.107387	0.008247	-0.114654
11	-1.153720	2.304289	1.353355	-0.659652	-0.219001	3.023613	-1.022502	0.531080	1.322670	-1.313569	...	-0.000985	0.169200	0.009480	0.042539	-0.069131

Reducir la dimensionalidad: El "reductor"

Reductor de Dimensionalidad (by Fluzo Corp.)

Reducir la dimensionalidad: El "reductor"

Reductor de Dimensionalidad (by Fluzo Corp.)

SELECCIÓN DE FEATURES:

- Mantiene features originales
-

Reducir la dimensionalidad: El "reductor"

Reductor de Dimensionalidad (by Fluzo Corp.)

SELECCIÓN DE FEATURES:

- Mantiene features originales
- Mecanismos Supervisados y no supervisados
-

Reducir la dimensionalidad: El "reductor"

Reductor de Dimensionalidad (by Fluzo Corp.)

SELECCIÓN DE FEATURES:

- Mantiene features originales
- Mecanismos Supervisados y no supervisados
- Perdida de relaciones con las features "eliminadas"

Reducir la dimensionalidad: El "reductor"

Reductor de Dimensionalidad (by Fluzo Corp.)

SELECCIÓN DE FEATURES:

- Mantiene features originales
- Mecanismos Supervisados y no supervisados
- Perdida de relaciones con las features "eliminadas"

TRANSFORMACION DE FEATURES:

- No mantiene las features originales
-

Reducir la dimensionalidad: El "reductor"

Reductor de Dimensionalidad (by Fluzo Corp.)

SELECCIÓN DE FEATURES:

- Mantiene features originales
- Mecanismos Supervisados y no supervisados
- Perdida de relaciones con las features "eliminadas"

TRANSFORMACION DE FEATURES:

- No mantiene las features originales
- El mecanismo más común es la "Proyección". Ej: PCA
- PCA permite "escoger" el grado de pérdida informativa

Reducir la dimensionalidad: El "reductor"

Reductor de Dimensionalidad (by Fluzo Corp.)

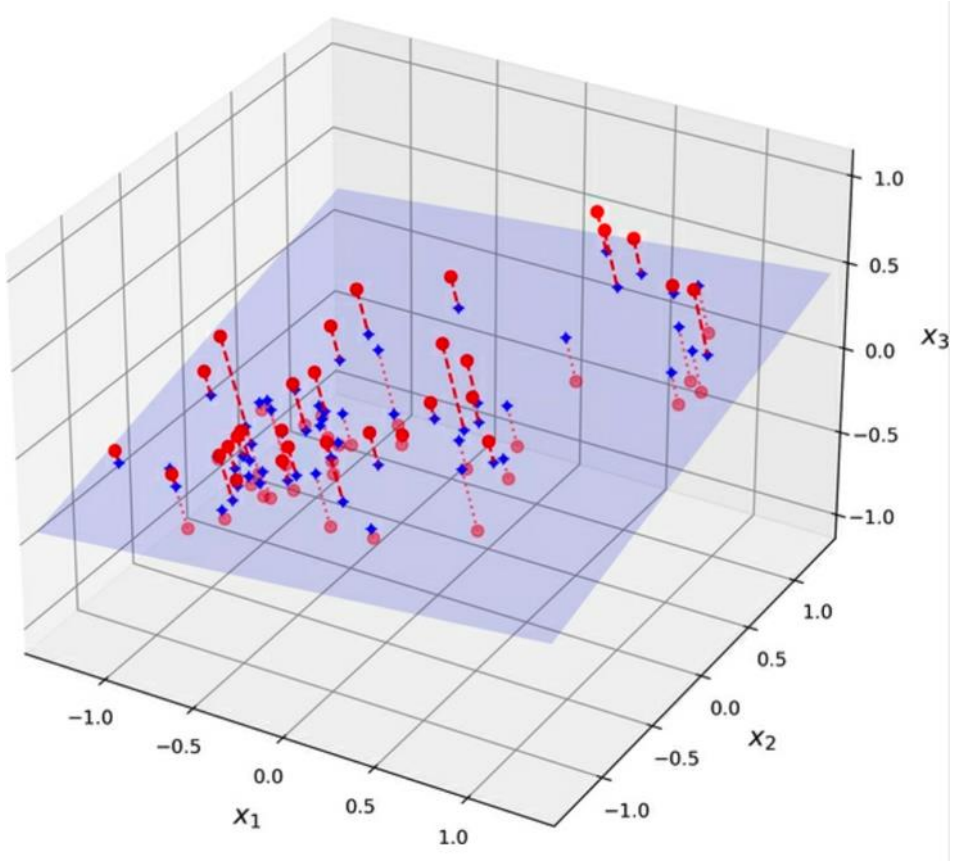
SELECCIÓN DE FEATURES:

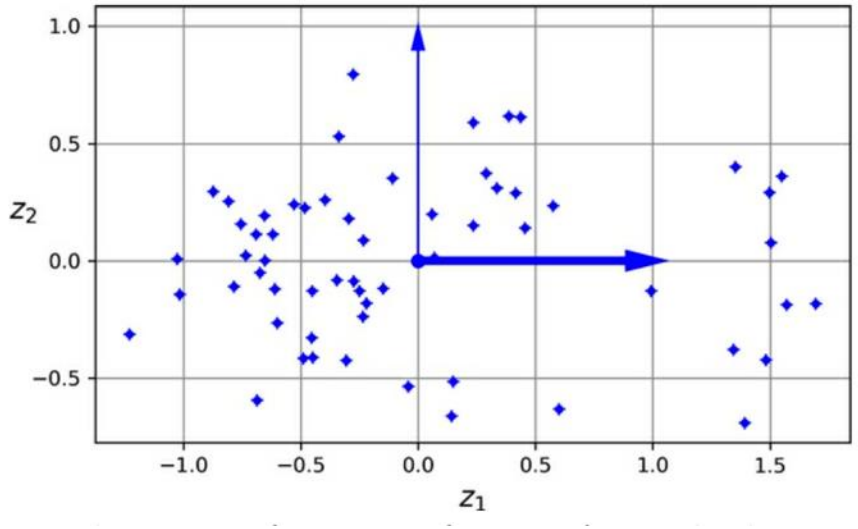
- Mantiene features originales
- Mecanismos Supervisados y no supervisados
- Perdida de relaciones con las features "eliminadas"

TRANSFORMACION DE FEATURES:

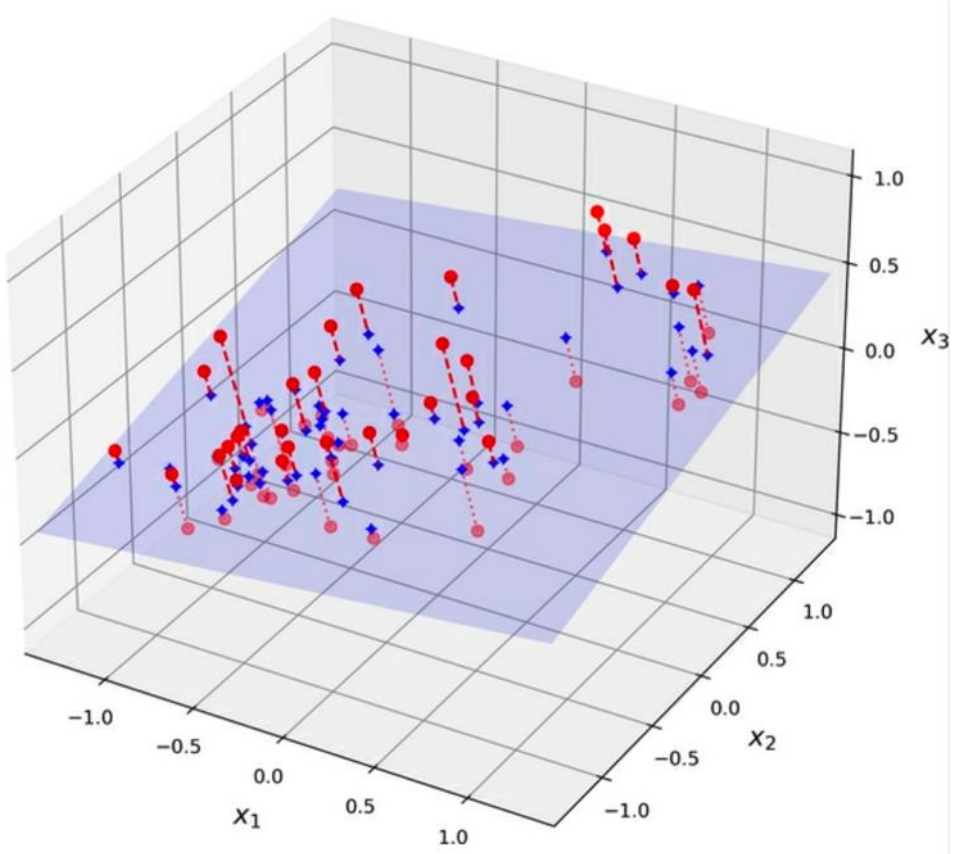
- No mantiene las features originales
- El mecanismo más común es la "Proyección". Ej: PCA
- PCA permite "escoger" el grado de pérdida informativa

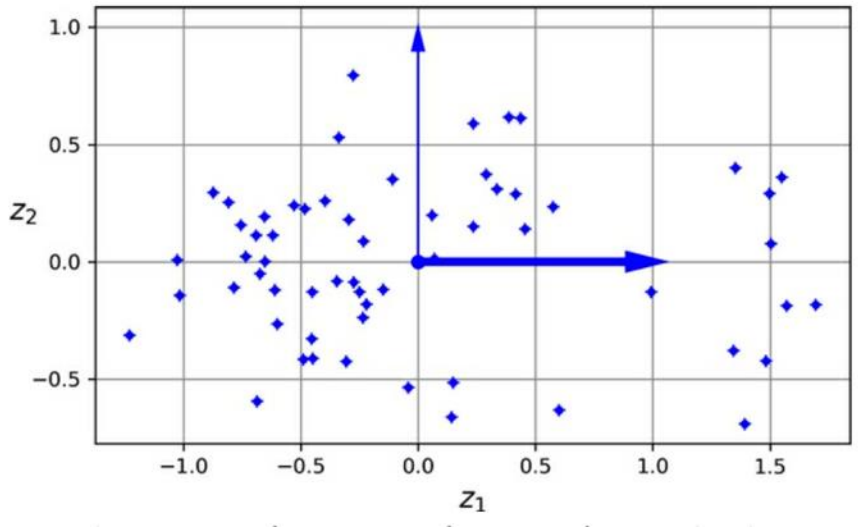
Reducir la dimensionalidad: Proyección





Reducir la dimensionalidad: Proyección





Un ejemplo simple

	x1	x2	x3
0	35.0	5.125890	-21.0
1	89.0	5.148074	15.0
2	61.0	5.000666	92.0
3	69.0	5.182370	-72.0
4	65.0	5.123824	51.0
5	92.0	5.164952	5.0

	x1	x3
0	35.0	-21.0
1	89.0	15.0
2	61.0	92.0
3	69.0	-72.0
4	65.0	51.0
5	92.0	5.0

Reducir la dimensionalidad


THE
BRIDGE