derivamos e igualamos a cero:

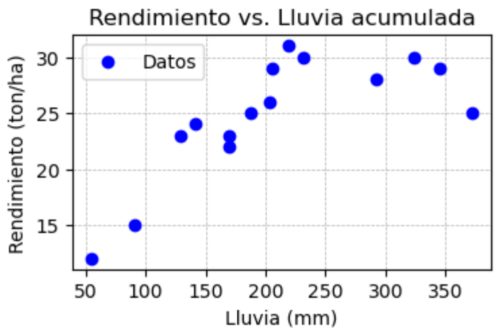
$$\begin{aligned}
\frac{\partial L_\alpha}{\partial \boldsymbol{w}} &= \frac{2}{N}\boldsymbol{Z}^\top \boldsymbol{Z}\boldsymbol{w} - \frac{2}{N}\boldsymbol{Z}^\top \boldsymbol{u} + 2\alpha\boldsymbol{w} = 0 \\
\left(\frac{1}{N}\boldsymbol{Z}^\top \boldsymbol{Z} + 2\alpha I\right)\boldsymbol{w} &= \frac{1}{N}\boldsymbol{Z}^\top \boldsymbol{u} \\
\widehat{\boldsymbol{w}}_\alpha &= \left(\operatorname{cov}\left(\left[\boldsymbol{x}^{(j)}\right]\right) + \alpha I\right)^{-1} \operatorname{cov}\left(\left[\boldsymbol{x}^{(j)}\right], \boldsymbol{y}\right)
\end{aligned}$$

Notar el efecto de α en acercar proporcionalmente la matriz de covarianzas a la identidad, lo cual mejora su invertibilidad.

12. Multicolinealidad en regresión polinomial

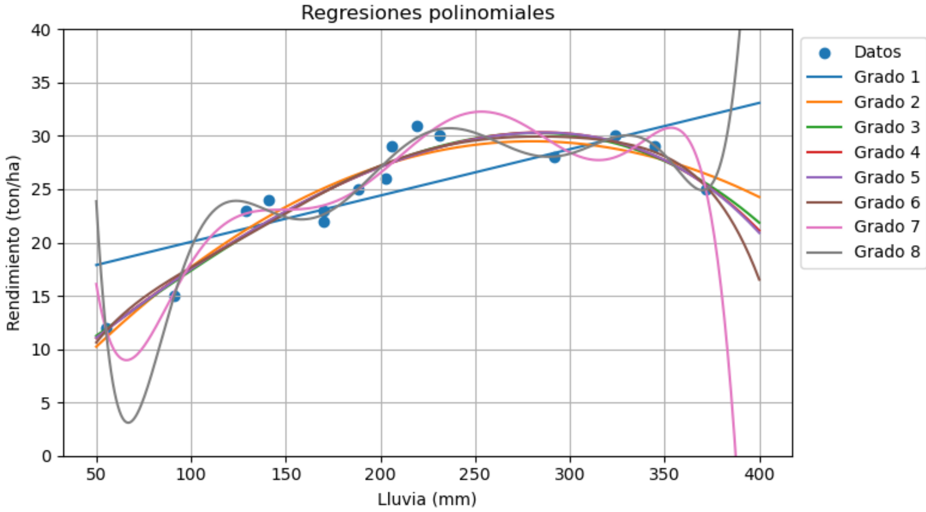
A modo de ilustración, considerar el siguiente conjunto de datos del rendimiento de un cultivo de papas en función de la lluvia acumulada:

x = Lluvia (mm)	y = Rendimiento (ton/ha)
206	29
188	25
219	31
372	25
345	29
231	30
203	26
170	23
55	12
91	15
292	28
141	24
129	23
170	22
324	30



Al mirar el gráfico vemos que la relación entre x e y no es lineal y parece razonable intentar con una regresión polinomial. Inmediatamente surge la pregunta de elegir el grado del polinomio.

El siguiente gráfico muestra varios polinomios, con grados que van desde 1 a 8, ajustados a estos datos:



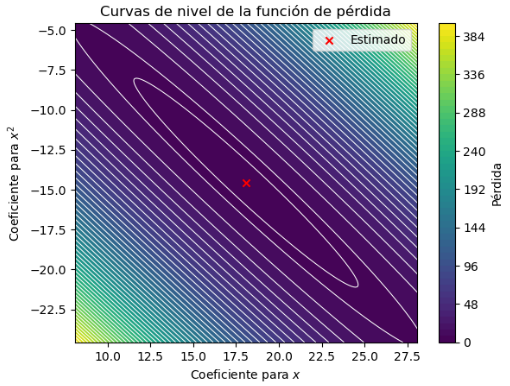
Para ajustar dichos polinomios siempre es conveniente estandarizar la matriz de diseño.

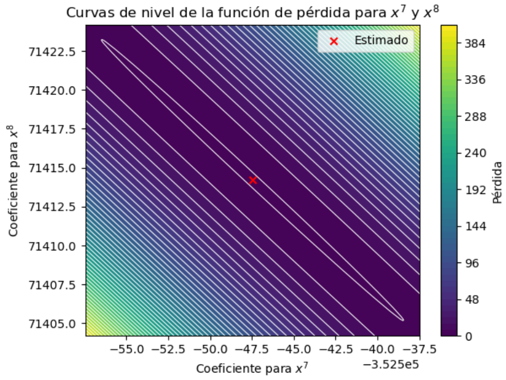
Observar la matriz de correlaciones para grado 8:

Cor	x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8
x^1	1	0.977	0.936	0.897	0.863	0.834	0.808	0.784
x^2	0.977	1	0.989	0.969	0.946	0.924	0.902	0.881
x^3	0.936	0.989	1	0.994	0.982	0.967	0.951	0.934
x^4	0.897	0.969	0.994	1	0.996	0.988	0.977	0.964
x^5	0.863	0.946	0.982	0.996	1	0.998	0.991	0.982
x^6	0.834	0.924	0.967	0.988	0.998	1	0.998	0.993
x^7	0.808	0.902	0.951	0.977	0.991	0.998	1	0.998
x^8	0.784	0.881	0.934	0.964	0.982	0.993	0.998	1

Más aún, el determinante de dicha matriz es $5{,}14 \times 10^{-39}$, y por lo tanto estamos ante la presencia de una marcada multicolinealidad, y por ende a riesgo de coeficientes grandes.

La multicolinealidad se puede visualizar al graficar las curvas de nivel de la función de pérdida. A modo de ejemplo:





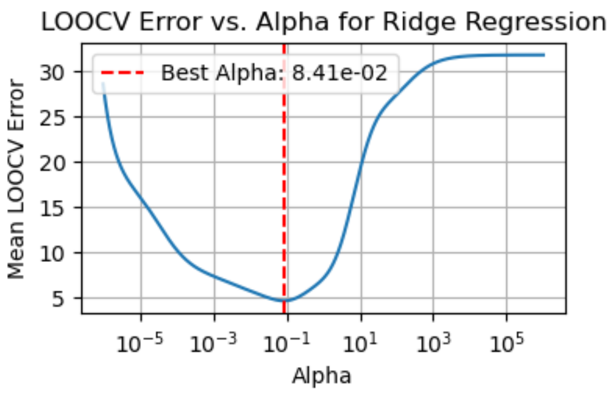
Las elipses estiradas (con los datos estandarizados) indican alta correlación entre los atributos.

Y de hecho los coeficientes para los distintos grados tienden a crecer rápidamente:

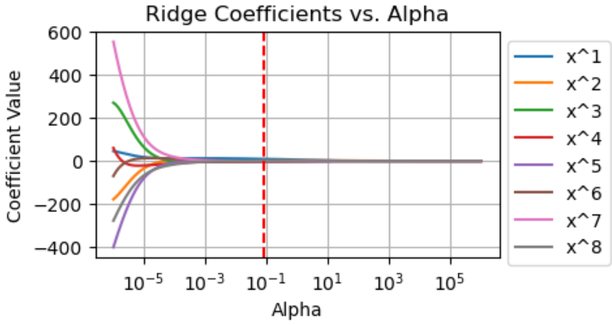
Coeficiente	**Grado 1**	**Grado 2**	**Grado 3**	**Grado 4**	**Grado 5**	**Grado 6**	**Grado 7**	**Grado 8**
w_1	3.85	18.08	10.79	15.02	12.71	76.59	-1533.75	-5491.59
w_2		-14.56	3.56	-13.02	-0.23	-453.11	13365.95	52990.47
w_3			-11.13	10.75	-16.30	1318.39	-50622.38	-231805.88
w_4				-9.49	15.63	-1950.95	102868.06	571182.95
w_5					-8.56	1419.35	-116379.46	-841240.02
w_6						-406.93	69088.87	735492.68
w_7							-16787.07	-352547.47
w_8								71414.21

Para controlar el tamaño de los coeficientes podemos correr una regresión polinomial de grado 8 regularizada. El problema se traslada ahora en elegir el valor de α.

El gráfico a continuación muestra la curva de error para varios valores de α, el error calculado usando la técnica de Leave One Out Cross Validation (LOOCV):



También podemos ver la evolución de los coeficientes en función de α:



Por último, podemos ver el efecto de la regularización en los coeficientes:

Coef	**Original**	**Ridge**
x^1	-5491.591405	9.290364281
x^2	52990.4687	1.239116016
x^3	-231805.8798	-2.013422183
x^4	571182.9466	-2.683297203
x^5	-841240.0155	-2.179430752
x^6	735492.6833	-1.218635966
x^7	-352547.4682	-0.16433541
x^8	71414.20573	0.79483848

Aquí tenemos el gráfico de la regresión regularizada: