

Modelo	Modelo matemático	Función de costo	Método de resolución	Correspondencia con el punto 1	Tamaño y observaciones
LinearRegressor	$y = Xw + \eta$	$J = \ Xw - y\ ^2$	Solución analítica mediante ecuaciones normales	Equivale a los mínimos cuadrados descritos en el punto 1	Rinde bien hasta del orden de 10^4 – 10^5 registros; muy veloz e interpretable.
Lasso	$y = Xw + \eta$	$J = \ Xw - y\ ^2 + \lambda\ w\ _1$	Descenso por coordenadas / gradiente proximal	Estimador MAP con prior Laplaciano sobre w	Favorece soluciones dispersas; útil para selección de variables.
ElasticNet	Igual	$J = \ Xw - y\ ^2 + \lambda_1\ w\ _1 + \lambda_2\ w\ ^2$	Descenso por coordenadas	Punto intermedio entre Lasso y Ridge	Suele funcionar mejor cuando las <i>features</i> están correlacionadas.
KernelRidge	$y = K\alpha + \eta$	$J = \ K\alpha - y\ ^2 + \lambda\alpha^\top K\alpha$	Solución cerrada en el problema dual	Ridge aplicado en el espacio de características $\phi(x)$	Coste cúbico $O(n^3)$; capacidad limitada en n .
SGDRegressor	Igual al lineal	$J = \ Xw - y\ ^2$	Optimización con gradiente estocástico	Aproxima la regresión lineal mediante SGD	Escala a conjuntos con millones de ejemplos.
BayesianRidge	$y = Xw + \eta$ con prior $w \sim \mathcal{N}(0, \lambda^{-1}I)$	Minimiza la evidencia marginal	Resolución cerrada; posterior Gaussiano	Versión Bayesiana (MAP/MLE) del modelo del punto 1	Entrega incertidumbre explícita sobre w .
GaussianProcessRegressor	$f(x) \sim \mathcal{GP}(0, k(x, x'))$	Maximiza la log-verosimilitud del kernel	Inversión/álgebra lineal de la matriz de kernel	Extiende KernelRidge a un enfoque no paramétrico	Complejidad $O(n^3)$; incertidumbre exacta.
SVR	$f(x) = w^\top \phi(x) + b$	$\frac{1}{2}\ w\ ^2 + C \sum \xi_i$, con margen ϵ	Programación cuadrática (SMO)	Regresión regularizada con kernel y margen ϵ	Escala de forma moderada; resistente a atípicos.
RandomForestRegressor	Ensamble de árboles	Minimiza el error cuadrático medio	<i>Bagging</i> con árboles entrenados en forma independiente	Modelo no lineal; sin vínculo directo con el punto 1	Muy escalable y robusto; no exige normalizar entradas.
GradientBoosting / XGBoost	Ensamble secuencial de árboles	Pérdida diferenciable optimizada con <i>boosting</i>	<i>Gradient boosting</i> iterativo (aditivo)	No lineal; combina aprendices débiles de manera secuencial	Alta precisión y buena escalabilidad; entrenamiento relativamente costoso.