

Formeln

Lineare Regression	Regularisierung	
<p>Hypothese:</p> $h_{\theta(x)} = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$ <p>Kostenfunktion (MSE):</p> $J(\theta) = \frac{1}{2n} \sum_{i=1}^n \left(h_{\theta(x^{(i)})} - y^{(i)} \right)^2$ <p>Ziel: $\min_{\theta} J(\theta)$</p> <p>Multivariat: Mehrere Features x_1, x_2, \dots, x_n</p> <p>Polynom-Regression:</p> $h_{\theta(x)} = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \dots$	<p>Kostenfunktion mit L2-Regularisierung:</p> $J(\theta) = \frac{1}{2n} \sum \left(h_{\theta(x^{(i)})} - y^{(i)} \right)^2 + \lambda \sum_{j=1}^d \theta_j^2$ <p>Effekt von λ:</p> <ul style="list-style-type: none">$\lambda = 0 \rightarrow$ kein Penaltygroßes $\lambda \rightarrow$ starke Bestrafung, Underfitting <p>Bias-Term θ_0 wird oft nicht regularisiert</p>	
Gradient Descent		
<p>Update-Regel:</p> $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$ <p>Für lineare Regression:</p> $\theta_j := \theta_j + \alpha \frac{1}{n} \sum_{i=1}^n \left(y^{(i)} - h_{\theta(x^{(i)})} \right) \cdot x_j^{(i)}$ <p>Lernrate α: Zu groß \rightarrow Divergenz, zu klein \rightarrow langsame Konvergenz</p>		
Logistische Regression	Support Vector Machines	
<p>Sigmoidfunktion:</p> $g(z) = \frac{1}{1 + e^{-z}}$ <p>Hypothese:</p> $h_{\theta(x)} = g(\theta^T x)$ <p>Klassifikation: $h_{\theta(x)} \geq 0.5 \rightarrow$ Klasse 1 $h_{\theta(x)} < 0.5 \rightarrow$ Klasse 0</p> <p>Entscheidungsgrenze:</p> $\theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0$ <p>Nicht-linearität durch Features wie $x_1^2, x_1 x_2, \dots$</p>	<p>Ziel:</p> $m_{w,b} \in \arg \min_{\{w,b\}} \left(\frac{1}{2} \ w\ ^2 + C \sum x_i \right)$ <p>Nebenbedingungen:</p> $y^{(i)} (w^T x^{(i)} + b) \geq 1 - \xi_i \text{ mit } \xi_i \geq 0$ <p>C kontrolliert Trade-off: großes $C \rightarrow$ weniger Fehler, kleines $C \rightarrow$ größerer Margin</p> <p>Kernel-Trick: z.B. $K(x, x') = e^{-\gamma \ x - x'\ ^2}$ (RBF-Kernel)</p>	