Model	Representation	Evaluation	Optimization
Logistic regression	$p(y=1\|\boldsymbol{x},\boldsymbol{w})=\sigma(\boldsymbol{w}^T\boldsymbol{x})$ $y=\begin{cases}1 & \text{if } p(y\|\boldsymbol{x},\boldsymbol{w})>0.5\\ 0 & \text{otherwise,}\end{cases}$	$\ell(\boldsymbol{w})=\prod_{i=1}^{N}p(y_i\|\boldsymbol{x_i},\boldsymbol{w})$ $NLL(\boldsymbol{w})=-\sum_{i=1}^{N}\Big[(y_i\log(\boldsymbol{w}^T\boldsymbol{x_i})+ (1-y_i)\log(1-\boldsymbol{w}^T\boldsymbol{x_i})\Big]$	Gradient descent (global minimum solution)
Neural Network	$h_{\boldsymbol{W},\boldsymbol{B}}(\boldsymbol{x})=f((\sum_{j=1}^{N_{L-1}}\boldsymbol{W}_{Lj}^{L-1}\boldsymbol{A}_j^{L-1}) + \boldsymbol{B}_0^{L-1})$ $\boldsymbol{A}_j^l=f(\sum_{j=1}^{N_{l-1}}\boldsymbol{W}_{lj}^{l-1}\boldsymbol{A}_j^{l-1}+\boldsymbol{B}_j^l)$	$J(\boldsymbol{W},\boldsymbol{B})=$ $\begin{cases}\frac{1}{2}\sum_{i=1}^{N}\left\|y_i-h_{\boldsymbol{W},\boldsymbol{B}}(\boldsymbol{x_i})\right\|_2^2 \text{ regression,}\\ \sum_{i=1}^{N}y_i\log(h_{\boldsymbol{W},\boldsymbol{B}}(\boldsymbol{x_i}))+\\ (1-y_i)\log(1-h_{\boldsymbol{W},\boldsymbol{B}}(\boldsymbol{x_i})) \text{ classification.}\end{cases}$	Backpropogation $\boldsymbol{W}_{ij}^l=\boldsymbol{W}_{ij}^l-\alpha\frac{\partial J(\boldsymbol{W},\boldsymbol{B})}{\partial \boldsymbol{W}_{ij}^l}$
Support vector machines	$f(\boldsymbol{x})=\begin{cases}1 & \boldsymbol{w}^T\boldsymbol{x_i}+b\geq 1-\xi_n\\ -1 & \boldsymbol{w}^T\boldsymbol{x_i}+b\leq -(1-\xi_n)\end{cases}$	$\min_{\boldsymbol{w},b,\xi}\ \frac{1}{2}\left\|\boldsymbol{w}\right\|_2^2+C\sum_{i=1}^{N}\xi_n$ s.t. $y_i(\boldsymbol{w}^Tx+b)\geq 1-\xi_n$, where $\xi=[\xi_1,\cdots,\xi_N]$, and $\xi_n\geq 0$.	Linear programming $L(\boldsymbol{w},b,\alpha,\xi)=\frac{1}{2}\left\|\boldsymbol{w}\right\|^2+C\sum_{i=1}^{N}\xi_n -\sum_{i=1}^{N}\alpha_n\Big[y_i(\boldsymbol{w}^T\boldsymbol{x}+b)-1+\xi_n\Big]+\sum_{i=1}^{N}\alpha_n$ where $\xi=[\xi_1,\cdots,\xi_N],\ \xi_i\geq 0$; $\alpha=[\alpha_1,\cdots,\alpha_N],\ \alpha_i\geq 0$.