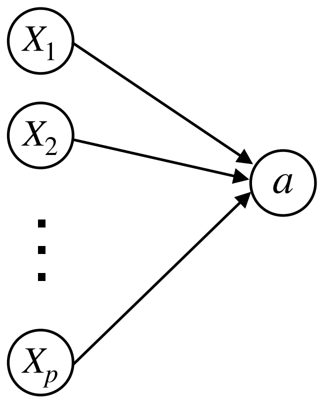
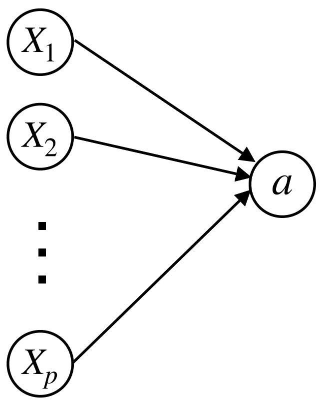
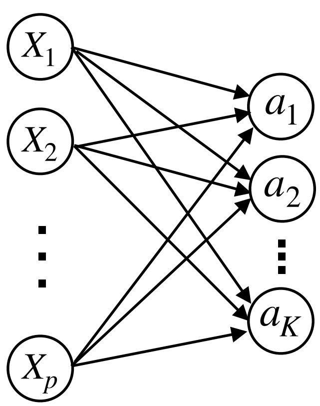
Model	$Y^{(i)}$	Model Diagram	Interpretation of $a^{(i)}$	Distribution for Response	Activation Function	Cost Function (negative log-likelihood)
Regression	A real number		Mean response for the given value of x	$Y^{(i)} \sim$ **Normal**$(a^{(i)}, \sigma^2)$	Identity/Linear $g(z) = z$	$J(b, w) = \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - a^{(i)}\right)^2$ The average of the squared residuals — note that this is equivalent to minimizing the sum of squared residuals or the root mean squared error.
Logistic Regression	0 or 1		Probability of class 1	$Y^{(i)} \sim$ **Bernoulli**$(a^{(i)})$	Sigmoid $g(z) = \sigma(z) = \frac{e^z}{1 + e^z}$	$J(b, w) = -\sum_{i=1}^{m} \left[y^{(i)} \log\left(a^{(i)}\right) + (1 - y^{(i)}) \log\left(1 - a^{(i)}\right)\right]$ Add up the log of the estimated probability for each observation's actual class, and multiply by -1.
Multinomial Regression	1, 2, …, K (K possible categories)		Vector of class probabilities: $a_j^{(i)}$ is the probability that $Y^{(i)}$ is in class j.	$Y^{(i)} \sim$ **Categorical**$(a^{(i)})$	Softmax $g(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$ For each class j = 1, …, K	$J(b, w) = -\sum_{i=1}^{m} \log\left(a_{y^{(i)}}^{(i)}\right)$ Add up the log of the estimated probability for each observation's actual class, and multiply by -1.