

$(X_{\text{raw}}, y_{X_{\text{raw}}})$

$X_{\text{raw}} = \text{"A gorgeous, witty, seductive movie."}$

Clean-up

"a gorgeous witty seductive movie"

Tokenisation

a gorgeous witty seductive movie

$X_{\text{token}} = [102, 12058, 33228, 10195, 39433, 1569]$

Text pre-processing

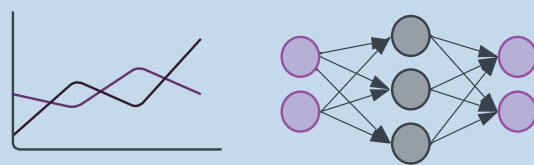
Tokens into
vectors

Feature extraction

token_id: 0 1 ... 102 ... 1569 ...

$X_{\text{one-hot}} = [[0, 0, \dots, 1, \dots],$
 $[0, 0, \dots, 0, \dots, 1, \dots] \dots]$

Feed into
the model



Model selection

Model evaluation

$\hat{y} = \operatorname{argmax}_y \Pr(y|X_{\text{one-hot}}; \theta)$

$\mathcal{L}_X = (\hat{y} - y_{X_{\text{raw}}})^2$

$\theta' = \theta - \eta \nabla \mathcal{L}_X$

$\mathbf{X}_{\text{test}} = [X'_1, X'_2, \dots, X'_n]$

Train, validate & test