

Recap of Basic Standard ML Doctrines

$\hat{y} = f_{\hat{\theta}}(x)$ $\hat{\theta}$ is learned params using loss and optimizers
 $\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \ell(f_{\theta}(x_i), y_i)$

↳ MLE \Rightarrow interpret loss as log-likelihood

Standard loss fns:

- squared error $\|y - \hat{y}\|^2$
 - logistic loss
 - hinge loss
 - cross-entropy \leftarrow multi-class
- ← binary

"overfitting":

Sols:

↳ Add regularizer $\left\{ \begin{array}{l} \rightarrow L_2 \text{ "ridge"} \\ \rightarrow L_1 \\ \rightarrow \text{etc.} \end{array} \right.$

↳ hyperparameters \rightarrow choose w/ validation

optimizer may have params associated with it:

↳ ex. learning rate in gradient descent

↳ sol: call it a hyperparameter, ignore it