* General
  + **m** Samples
  + **n** Features
  + **X** – design matrix (mXn)
  + **Y** – response/label, ground truth (mX1)
* Linear regression
  + – coefficients (nX1)
  + Model (Hypothesis class\function family) – linear functions: ()
  + Fit – finding the optimal
  + Learning principle (How do we pick a specific function – the coefficients, from the model) – OLS:
    - Criteria (how we got the closed above solution) – MSE:
  + Prediction – applying in the model to find the prediction -
  + Interpretability - With a fitted model, each coefficient denotes the change in the predicted outcome for a one-unit change in the corresponding feature.
* Logistic regression
  + Distribution Family – Bernoulli:
  + Model (Hypothesis class\function family) – linear functions + a composite sigmoid function: , ()
  + Fit – finding the optimal – )in our case indirectly p(
  + Learning principle – MLE:
    - Criteria – maximizing log-likelihood: – no closed form solution, solve with Gradient Descent
  + Prediction – specify a thresholding parameter and predict
  + Interpretability - With a fitted model, each coefficient denotes the change in the predicted outcome for a one-unit change in the corresponding feature.