COMP90051

Workshop Week 03

About the Workshops

- 7 sessions in total
  - Tue 12:00-13:00 AH211
  - Tue 12:00-13:00 AH108 *
  - Tue 13:00-14:00 AH210
  - Tue 16:15-17:15 AH109
  - Tue 17:15-18:15 AH236 *
  - Tue 18:15-19:15 AH236 *
  - Fri 14:15-15:15 AH211

About the Workshops

- Homepage
  - https://trevorcohn.github.io/comp90051-2017/workshops

- Solutions have been released.

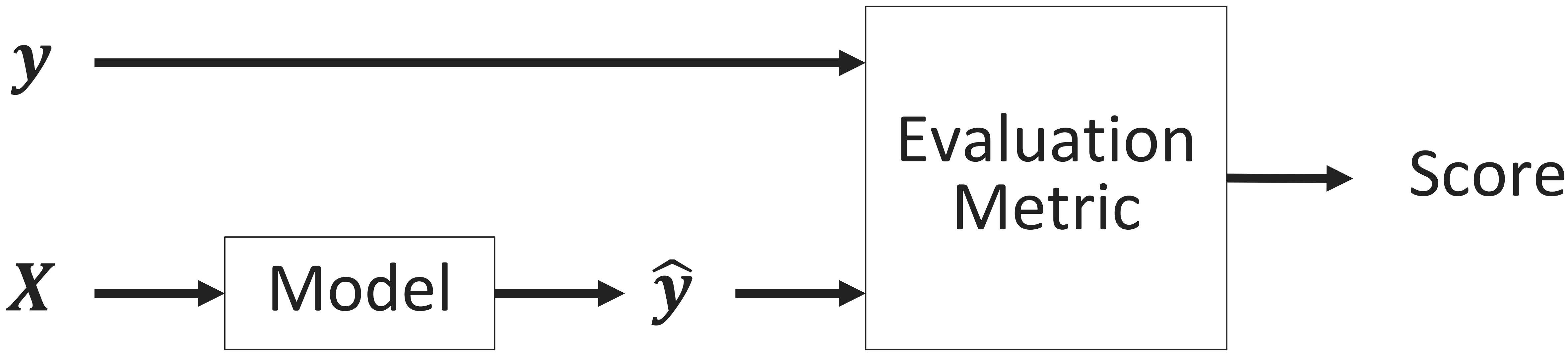
Outline

- Review the lecture, background knowledge, etc.
  - Model evaluation, selection, optimization
  - Regularizor as a prior
- Jupyter Usage
- Notebook tasks
  - Task 1: Linear regression
  - Task 2: Polynominal regression

Outline

- Review the lecture, background knowledge, etc.
  - Model evaluation, selection, optimization
  - Regularizor as a prior
- Jupyter Usage
- Notebook tasks
  - Task 1: Linear regression
  - Task 2: Polynominal regression

Model Evaluation

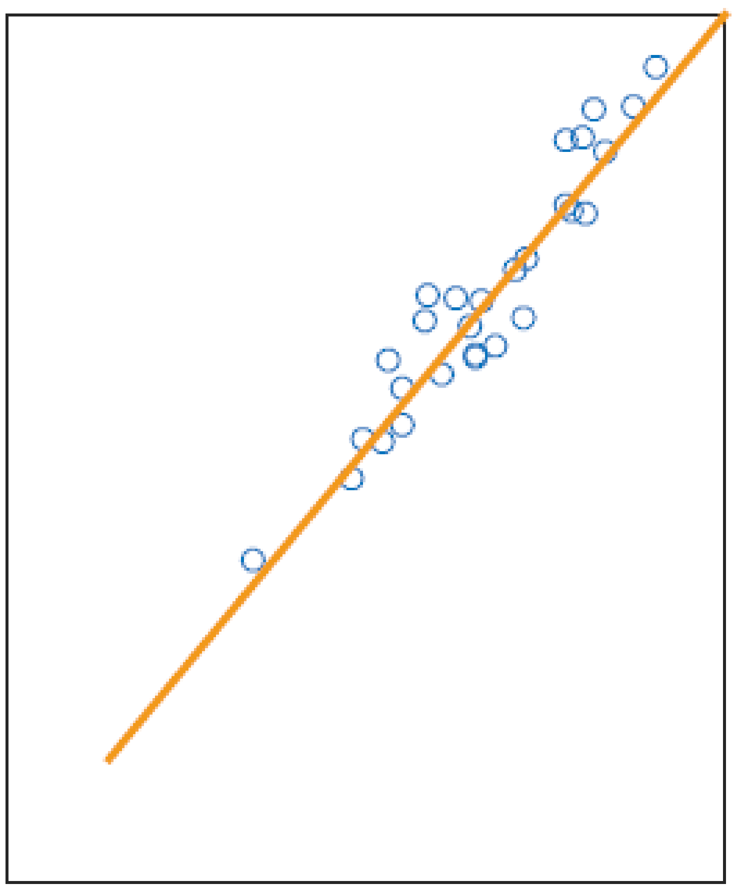


What could the output be?

- Regression
  - A value
  - A distribution

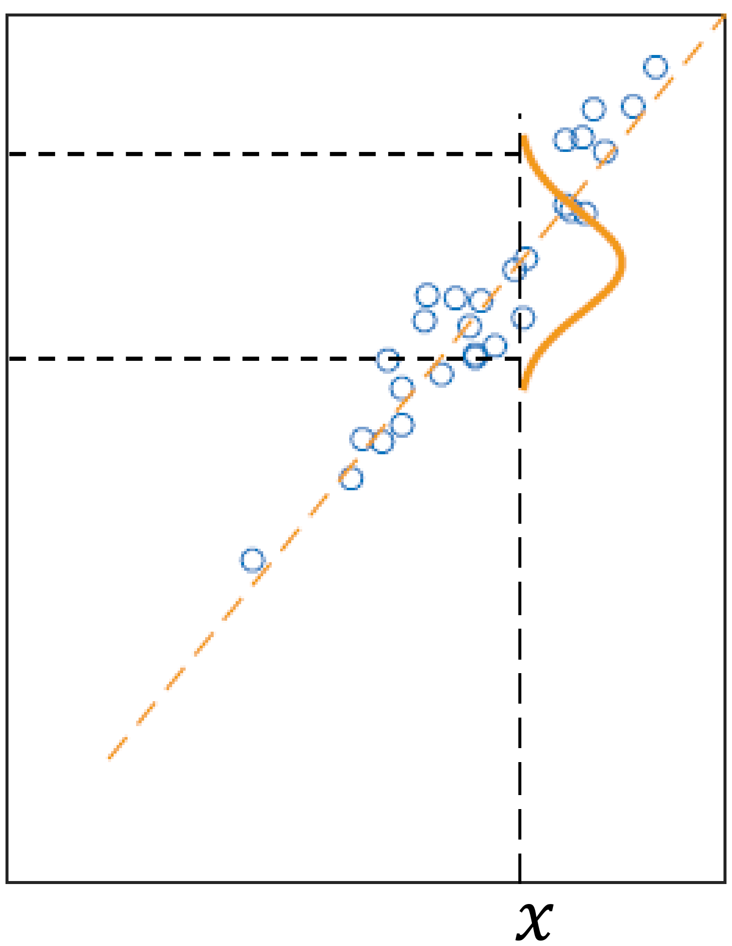
- Classification
  - A label
  - A value (binary) / values (multi-class)
  - A distribution

Types of models



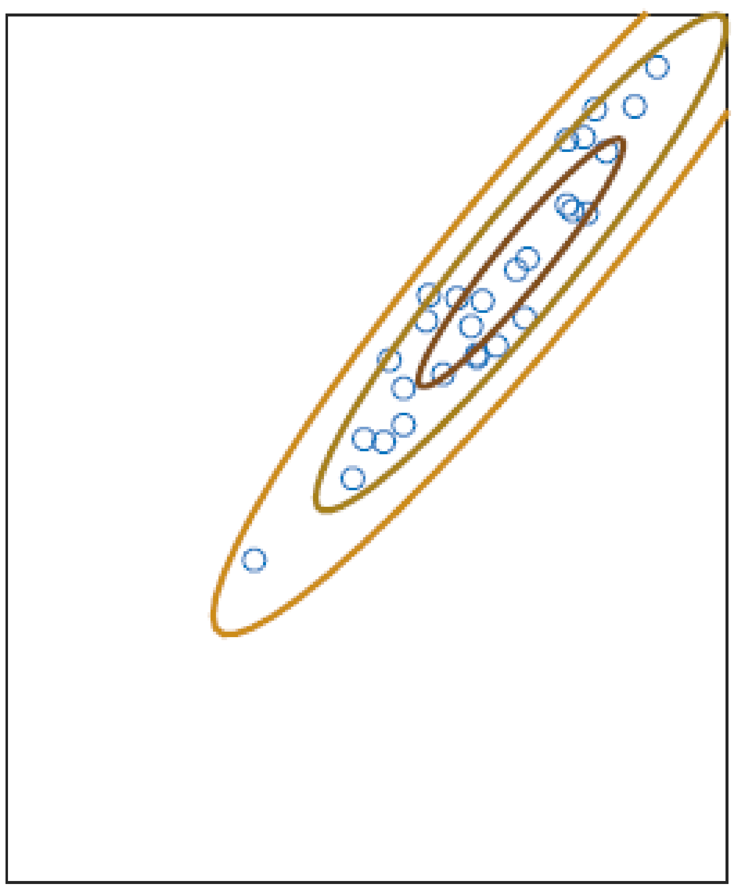
$\hat{y} = f(x)$

KT mark was 95, ML mark is predicted to be 95



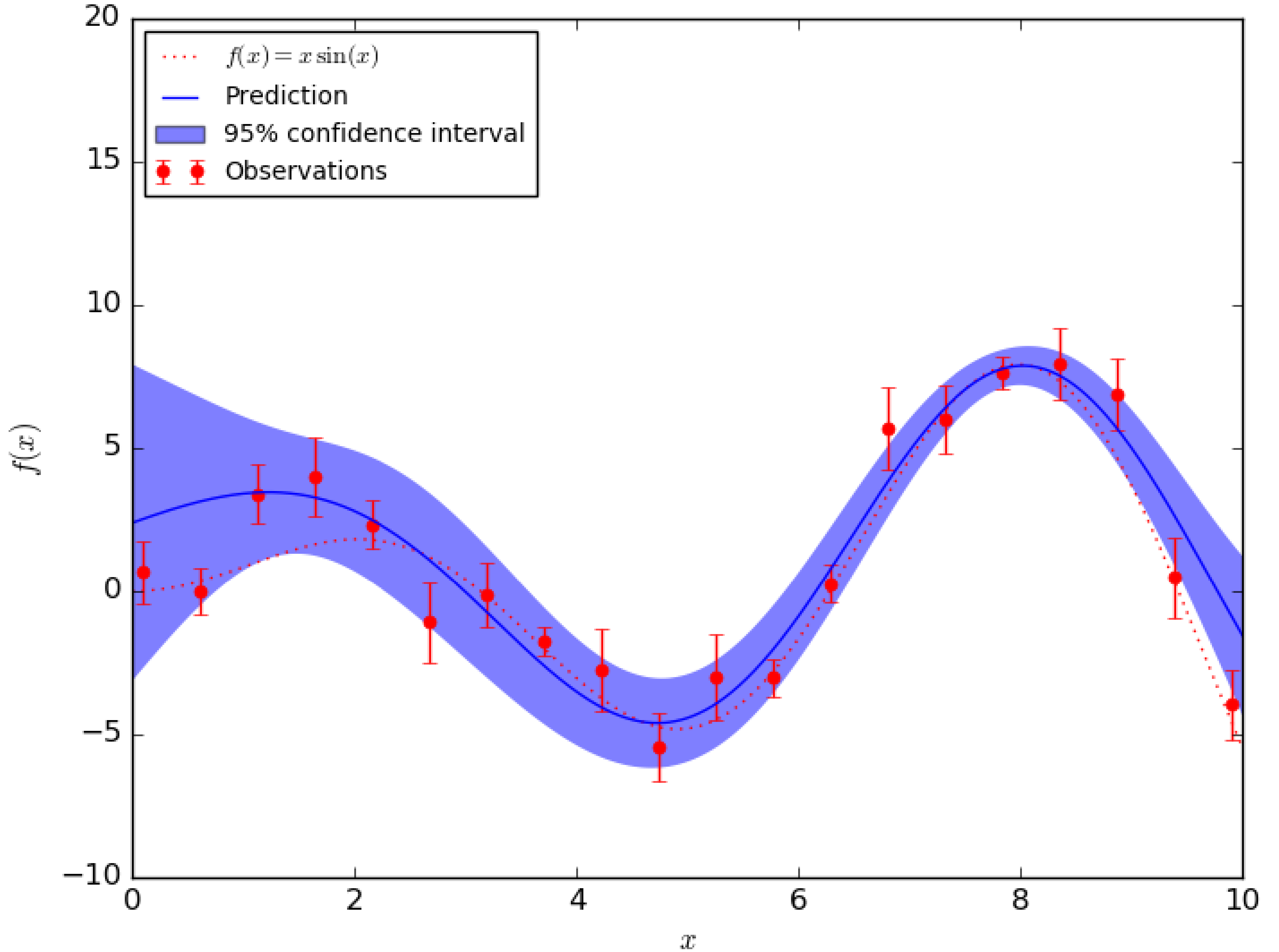
$P(y|x)$

KT mark was 95, ML mark is likely to be in (92, 97)

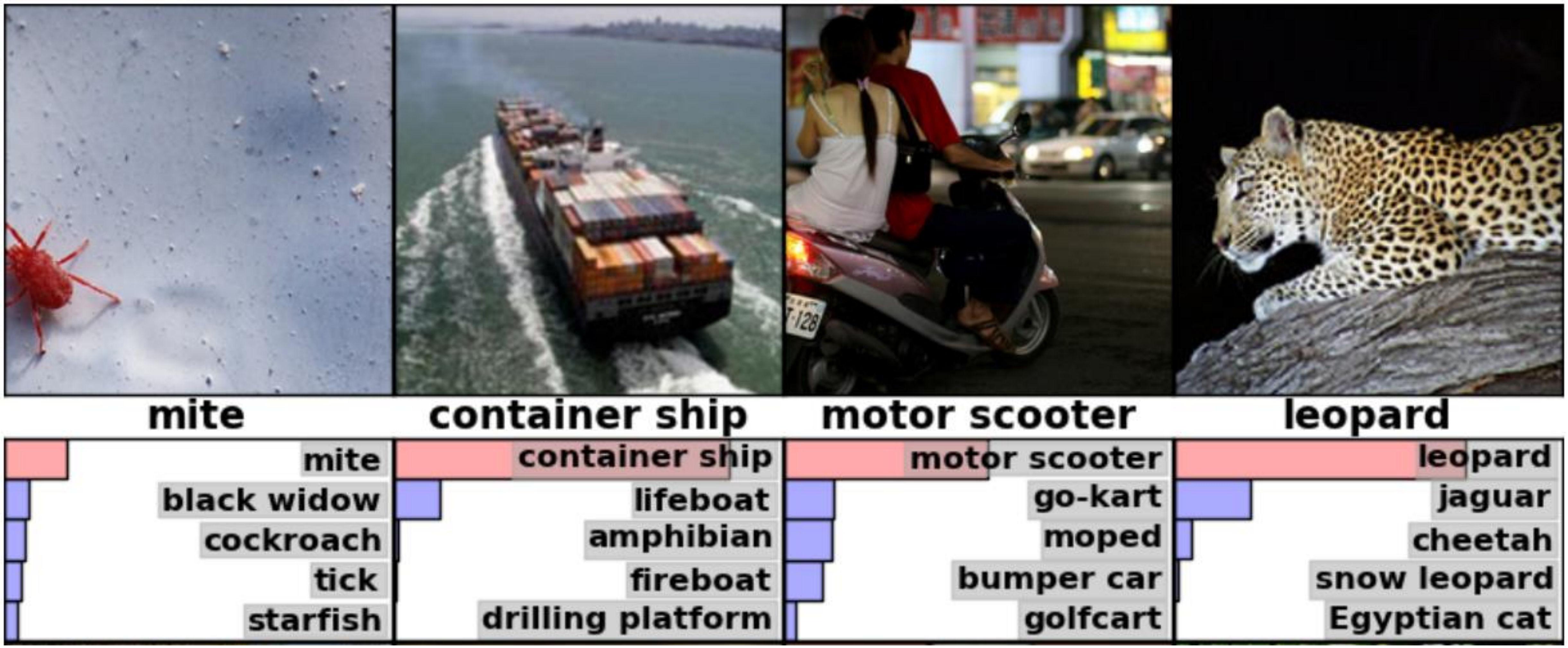


$P(x, y)$

probability of having $(KT = x, ML = y)$



http://scikit-learn.org/0.17/auto_examples/gaussian_process/plot_gp_regression.html



https://www.tensorflow.org/tutorials/image_recognition

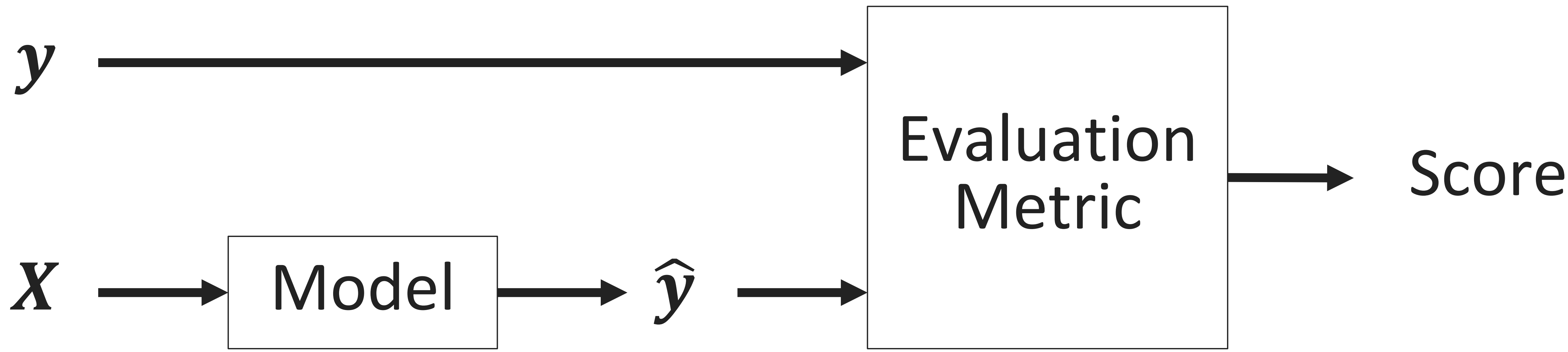
sklearn.linear_model.LogisticRegression

Methods

decision_function (X)	Predict confidence scores for samples.
densify ()	Convert coefficient matrix to dense array format.
fit (X, y[, sample_weight])	Fit the model according to the given training data.
fit_transform (X[, y])	Fit to data, then transform it.
get_params ([deep])	Get parameters for this estimator.
predict (X)	Predict class labels for samples in X.
predict_log_proba (X)	Log of probability estimates.
predict_proba (X)	Probability estimates.
score (X, y[, sample_weight])	Returns the mean accuracy on the given test data and labels.

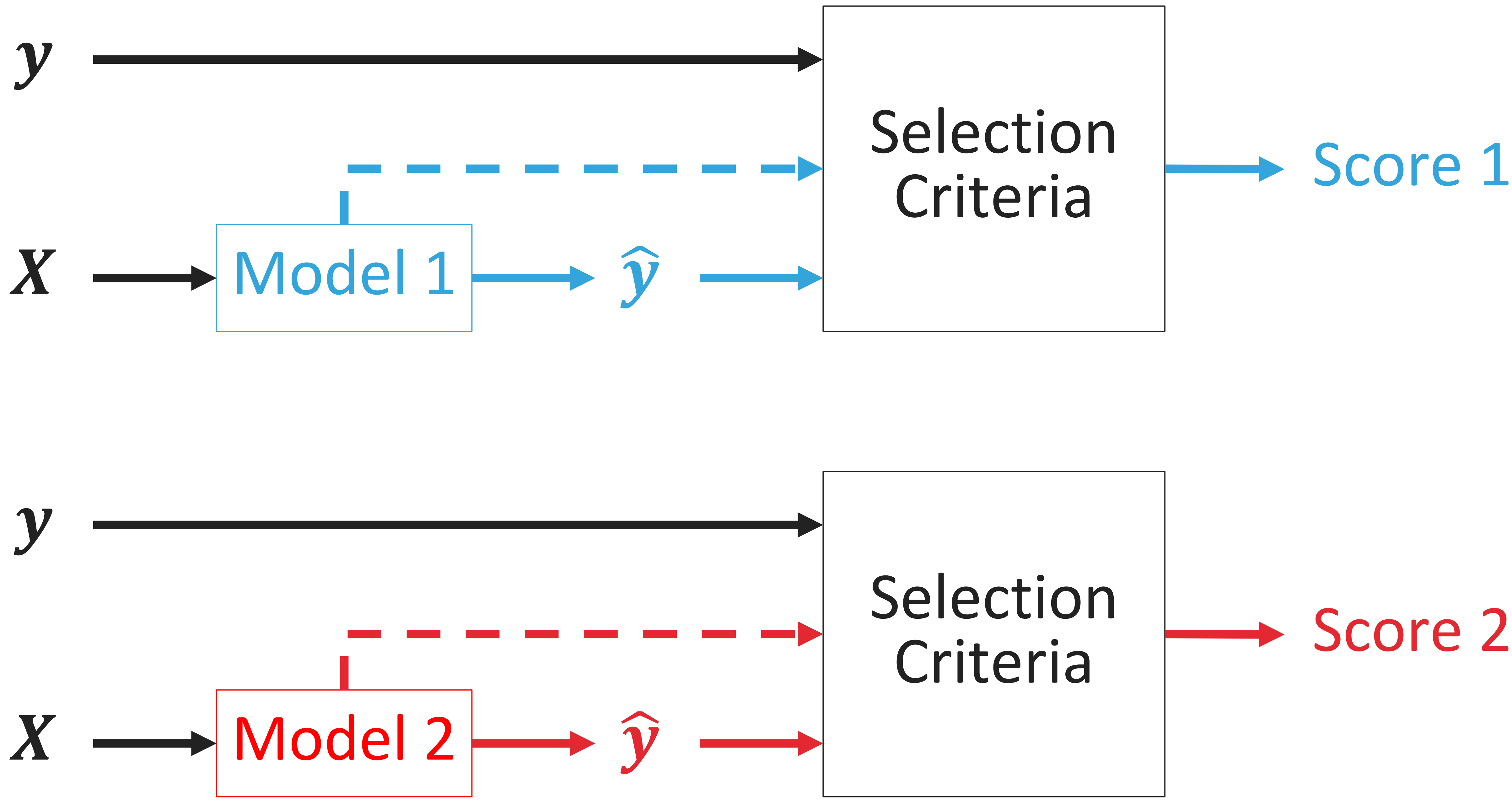
http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Model Evaluation

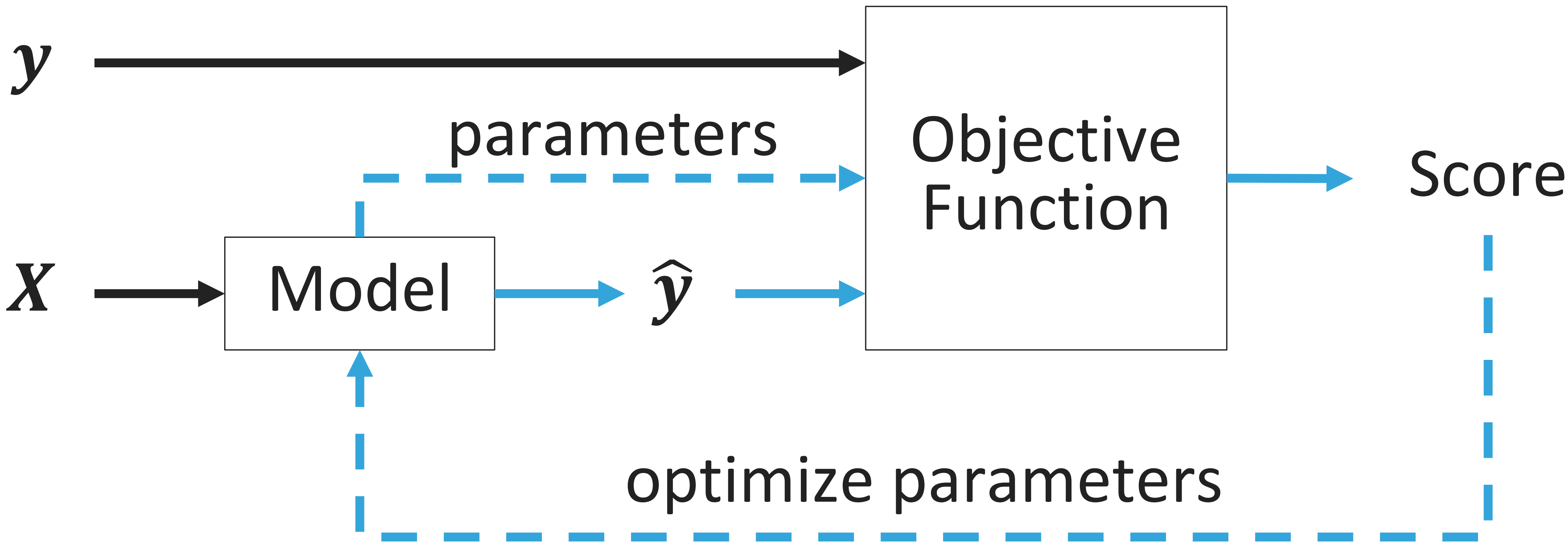


- Regression
  - RMSE, MAE, etc.
- Classification
  - Accuracy, precision, recall, f-score, etc.
  - Log-loss (a.k.a. cross entropy), likelihood, etc.

Model Selection



Model Optimization



- The evaluation metric & the objective function may differ
  - Could be entirely different
  - Or additional terms in the objective function, e.g. L1/L2

More on the objective function

- Maximize the likelihood (or log likelihood)

$$\max_{\boldsymbol{w}} p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{w}) \;\Leftrightarrow\; \max_{\boldsymbol{w}} \prod p(y_i|\boldsymbol{x}_i,\boldsymbol{w}) \;\Leftrightarrow\; \max_{\boldsymbol{w}} \sum \log p(y_i|\boldsymbol{x}_i,\boldsymbol{w})$$

- Maximize the posterior (a.k.a. max a posteriori, MAP)

$$\max_{\boldsymbol{w}} p(\boldsymbol{w}|\boldsymbol{X},\boldsymbol{y}) \;\rightarrow\; \max_{\boldsymbol{w}} p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{w})p(\boldsymbol{w}) \quad (\text{assume } \boldsymbol{w} \perp \boldsymbol{X})$$

- Minimize the loss function (+regularization)

$$\min_{\boldsymbol{w}} \sum L(f(\boldsymbol{x}_i;\boldsymbol{w}),y_i) \quad \text{or} \quad \min_{\boldsymbol{w}} \sum L(f(\boldsymbol{x}_i;\boldsymbol{w}),y_i) + \lambda R(\boldsymbol{w})$$

  - Minimize the log-loss (a.k.a. cross entropy) (+L1/L2)
  - Minimize the hinge-loss (+L2)
  - Minimize the mean squared error (+L1/L2)

More on the objective function

- Maximize the likelihood (or log likelihood)

$$\max_{\boldsymbol{w}} p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{w}) \Leftrightarrow \max_{\boldsymbol{w}} \prod p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) \Leftrightarrow \max_{\boldsymbol{w}} \sum \log p(y_i|\boldsymbol{x}_i, \boldsymbol{w})$$

- Maximize the posterior (a.k.a. max a posteriori, MAP)

$$\max_{\boldsymbol{w}} p(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{y}) \quad \rightarrow \quad \max_{\boldsymbol{w}} p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{w}) p(\boldsymbol{w}) \quad \text{(assume } \boldsymbol{w} \perp \boldsymbol{X}\text{)}$$

- Minimize the loss function (+regularization)

$$\min_{\boldsymbol{w}} \sum L(f(\boldsymbol{x}_i; \boldsymbol{w}), y_i) \quad \text{or} \quad \min_{\boldsymbol{w}} \sum L(f(\boldsymbol{x}_i; \boldsymbol{w}), y_i) + \lambda R(\boldsymbol{w})$$

  - Minimize the log-loss (a.k.a. cross entropy) (+L1/L2)
  - Minimize the hinge-loss (+L2)
  - Minimize the mean squared error (+L1/L2)

Outline

- Review the lecture, background knowledge, etc.
  - Model evaluation, selection, optimization
  - Regularizor as a prior
- Jupyter Usage
- Notebook tasks
  - Task 1: Linear regression
  - Task 2: Polynominal regression

Regulariser as a prior

- Without regularisation model parameters are found based entirely on the information contained in the training set $\boldsymbol{X}$
- Regularisation essentially means introducing additional information
- Recall our probabilistic model $\mathcal{Y} = \boldsymbol{x}'\boldsymbol{w} + \varepsilon$
  * Here $\mathcal{Y}$ and ε are random variables, where ε denotes noise
- Now suppose that $\boldsymbol{w}$ is also a random variable (denoted as $\mathcal{W}$) with a normal prior distribution

$$\mathcal{W} \sim \mathcal{N}(0, \lambda^2)$$

Computing posterior using Bayes rule

- The prior is then used to compute the posterior

posterior likelihood prior

$$p(\boldsymbol{w}|\boldsymbol{X},\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{y}|\boldsymbol{X})}$$

marginal likelihood

- Instead of maximum likelihood (MLE), take ***maximum a posteriori*** estimate (MAP)
- Apply log trick, so that

$$\log(posterior) = \log(likelihood) + \log(prior) - \cancel{\log(marg)}$$

- Arrive at the problem of minimising

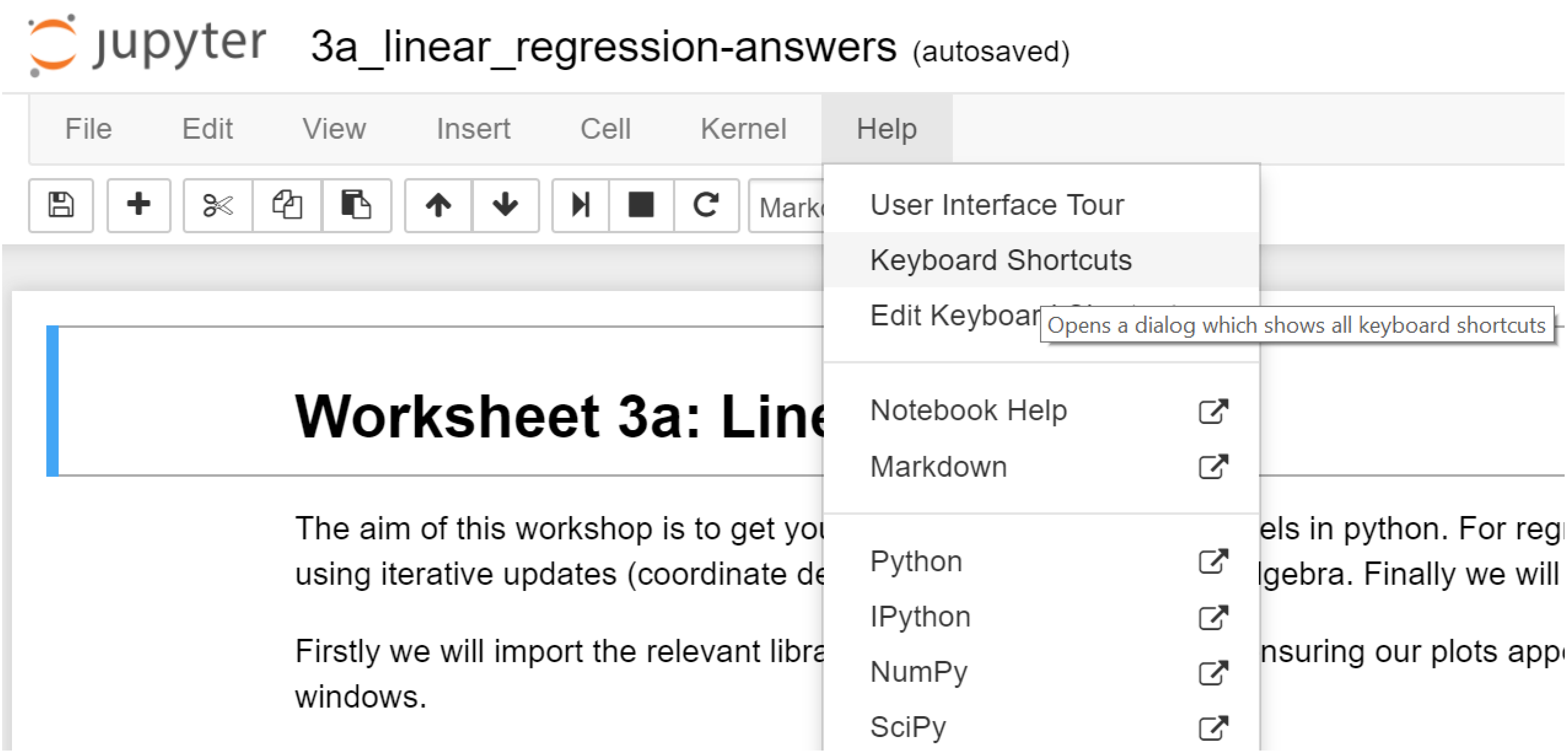
$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \lambda\|\boldsymbol{w}\|_2^2$$

this term doesn't affect optimisation

Outline

- Review the lecture, background knowledge, etc.
  - Model evaluation, selection, optimization
  - Regularizor as a prior
- Jupyter Usage
- Notebook tasks
  - Task 1: Linear regression
  - Task 2: Polynominal regression

Keyboard Shortcuts



Outline

- Review the lecture, background knowledge, etc.
  - Model evaluation, selection, optimization
  - Regularizor as a prior
- Jupyter Usage
- Notebook tasks
  - Task 1: Linear regression
  - Task 2: Polynomial regression

Linear regression

- $x_1 \rightarrow y_1, x_2 \rightarrow y_2, x_3 \rightarrow y_3, x_4 \rightarrow y_4$
- $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{bmatrix}$
- Minimize the objective function

$$\frac{1}{4}\sum_{i=1}^{4}(\hat{y}_i - y_i)^2 \quad \text{or} \quad \frac{1}{4}\sum_{i=1}^{4}(\hat{y}_i - y_i)^2 + \lambda \sum_{j=0}^{1} w_j^2$$

- Analytic solution & iterative solution

Multivariate linear regression (2-D points)

- $(x_{1,1}, x_{1,2}) \rightarrow y_1, (x_{2,1}, x_{2,2}) \rightarrow y_2$
- $(x_{3,1}, x_{3,2}) \rightarrow y_3, (x_{4,1}, x_{4,2}) \rightarrow y_4$
- $\begin{bmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ 1 & x_{3,1} & x_{3,2} \\ 1 & x_{4,1} & x_{4,2} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{bmatrix}$
- Minimize the objective function

$$\frac{1}{4}\sum_{i=1}^{4}(\hat{y}_i - y_i)^2 \quad \text{or} \quad \frac{1}{4}\sum_{i=1}^{4}(\hat{y}_i - y_i)^2 + \lambda \sum_{j=0}^{2} w_j^2$$

- Analytic solution & iterative solution

Polynomial regression (Quadratic)

- $x_1 \to y_1, x_2 \to y_2, x_3 \to y_3, x_4 \to y_4$
- $\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{bmatrix}$
- Minimize the objective function

$$\frac{1}{4}\sum_{i=1}^{4}(\hat{y}_i - y_i)^2 \quad \text{or} \quad \frac{1}{4}\sum_{i=1}^{4}(\hat{y}_i - y_i)^2 + \lambda \sum_{j=0}^{2} w_j^2$$

- Analytic solution & iterative solution