

## Assignment-1 (CS-419)

1. Naive-Bayes classifier with discrete input features i.e., categorical features. In terms of training and test set error, compare solutions of
  - (a) MLE (using multivariate multinoulli naive-Bayes model; sec. 3.5, 3.5.1.1)
  - (b) MAP (using multivariate dirichlet-multinoulli naive-Bayes model; sec 3.5.1.2). Assume reasonable hyper-parameters.
  - (c) Bayesian naive-Bayes classifier (sec 3.5.2)
2. Bayes classifier with real-valued input features. In terms of training and test set error, compare solutions of
  - (a) MLE (using multivariate Normal model; sec. 4.1.3, 4.2; equivalently 'Gaussian Discriminative Analysis')
  - (b) MAP (using NIW model; sec. 4.6.3, 4.6.3.4). Assume reasonable hyper-parameters.
  - (c) Bayesian Bayes classifier (sec. 4.6.3.6)
3. Binary classification using discriminative approaches. In terms of training and test set error, compare solutions of
  - (a) Logistic Regression<sup>1</sup>. Perform 10-fold cross-validation to choose the hyper-parameter.
  - (b) Using the MAP estimate from above, obtain the Gaussian approximation of the posterior and now perform Bayesian logistic regression on the dataset using Monte-Carlo approximation (sec. 8.4.4.1).
  - (c) Using the MAP estimate from above, obtain the Gaussian approximation of the posterior and now perform Bayesian logistic regression on the dataset using Probit approximation (sec. 8.4.4.2).
4. Binary classification using frequentist approaches. In terms of training and test set error, compare solutions with
  - (a) hinge-loss (soft-margin L1-Loss SVM<sup>2</sup>)
  - (b) squared-hinge-loss (soft-margin L2-Loss SVM<sup>3</sup>)

---

<sup>1</sup>Logistic regression code available at <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

<sup>2</sup>Code available at <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

<sup>3</sup>Code available at <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

- (c) linear-loss ( $l(y, w^\top x) = -yw^\top x$ ). This is a simple unconstrained convex quadratic program.

In all cases choose the hyper-parameter using 10-fold cross-validation.

5. Linear regression. In terms of training and test set error, compare solutions with:

- (a) Least-square regression
- (b) Ridge-regression
- (c) Support Vector Regression (L1-Loss Linear SVR<sup>4</sup>)

Wherever required, perform 10-fold cross-validation to choose the best hyper-parameter.

Your dataset can be downloaded from <http://archive.ics.uci.edu/ml/datasets.html>.

---

<sup>4</sup>Code available at <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>