A distributed algorithm for fitting generalized additive models

CS 584: Big Data Analytics

Generalized Linear Models (GLMs)

- Linear predictor

$$\eta(x) = \beta_0 + x\beta$$

- Two functions:
  - Link function describes how mean depends on the linear predictor

$$\eta(x) = g(r(x))$$

  - Variance function that describes how variance depends on the mean

$$\mathrm{Var}[Y|X = x] = \sigma^2 V(r(x))$$

Example: Logistic Regression as GLM

- Logistic regression probability

$$P(Y = y|x) = \frac{1}{1 + \exp(-y(\beta x + \beta_0))}$$

- Link Function: Logit link

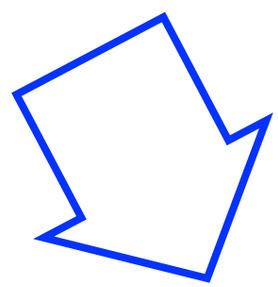
$$\eta = \text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right)$$

- Variance function

$$V(r) = r(1 - r), \sigma^2 = 1$$

Linear models to Generalized Additive Model

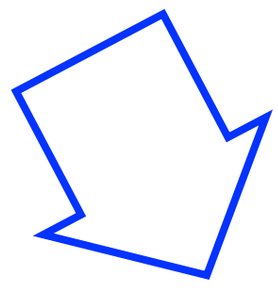
$$y = \beta_0 + x\beta + \epsilon$$



link function relates the conditional mean to the linear predictor

$$\eta = \beta_0 + x\beta$$

$$E[y] = g(\mu) = \eta$$



no longer require a linear predictor

$$g(\mu) = \beta_0 + \sum_j f_j(x_j)$$

Generalized Additive Model (GAM)

- Local regression to low dimensional projections of the data

- Surface estimation via a collection of one-dimensional functions (each function is analogous to coefficients in linear regression)

- Ability to model nonlinearities in the data automatically

Generalized Additive Models

- Optimization problem formulation

$$\min \; L\left(\sum_{i=1}^{N} f_i(x_i; p_i)\right) + \sum_{i=1}^{N} r_i(p_i)$$

- L is a loss function appropriately chosen for the link function
- pi are the parameters that specify each function fi
- Regularization is based on the parameters

ADMM Algorithm

- Introduce dummy variables to obtain a sharing formulation:

$$\min \; L\left(\sum_{i=1}^{N} z_i\right) + \sum_{i=1}^{N} r_i(p_i)$$

$$\text{s.t. } z_i = f_i(x_i; p_i)$$

- Update can be derived using sharing updates:

$$p_i^{k+1} = \text{argmin } r_i(p_i) + \rho/2||f_i(x_i; p_i) - f_i(x_i; p_i^k) + \bar{f}^k - \bar{z}^k + u^k||_2^2$$

$$\bar{z}^{k+1} = \text{argmin}_{\bar{z}} \; L(N\bar{z}) + (N\rho/2)||\bar{z} - u^k - \bar{f}^{k+1}||_2^2$$

$$u^{k+1} = u^k + \bar{f}^{k+1} - \bar{z}^{k+1}$$

ADMM Algorithm

- Each pi update fits a function that minimizes the total square error of the function evaluated at the data points in the vector xi

- Each pi update is independent so can be carried in parallel

- z update only evaluates the fitted functions at xi

- Algorithm can be run in parallel rather compared to backfitting (traditional algorithm) which is sequential

Example: Spam Data

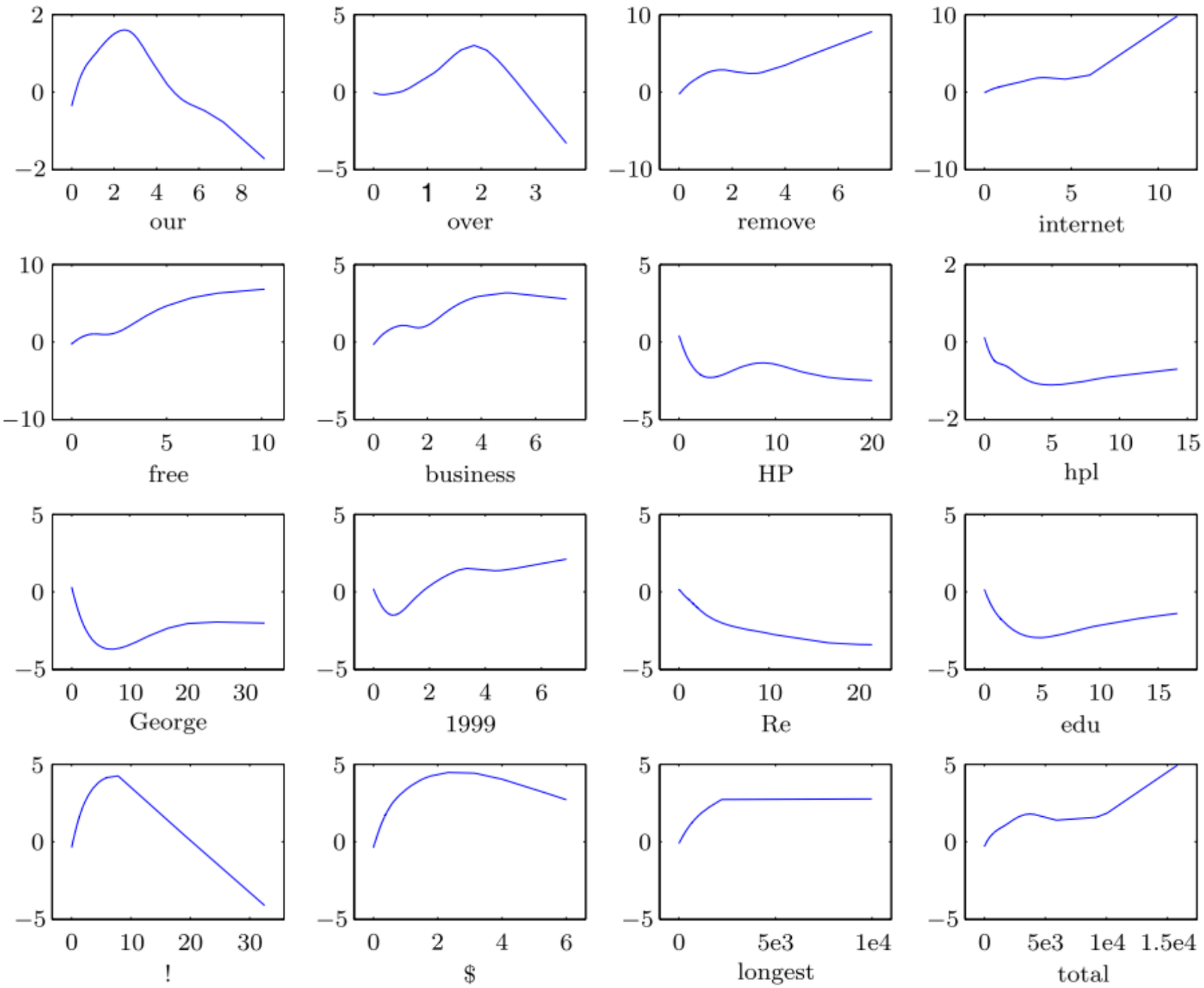
- Classify whether a sample is an email or spam

- 57 features or predictors based on frequency of word appearances, average length, longest length, and sum of all length of uninterrupted sequence of capital letters

- Use an additive logistic model

$$\text{Prob}(y_j = 1) = \frac{\exp(\sum(f_i(x_i; p_i))_j)}{1 + \exp(\sum(f_i(x_i; p_i))_j)}$$

Example: Spam Functions



Example: Coronary Heart Disease Prediction

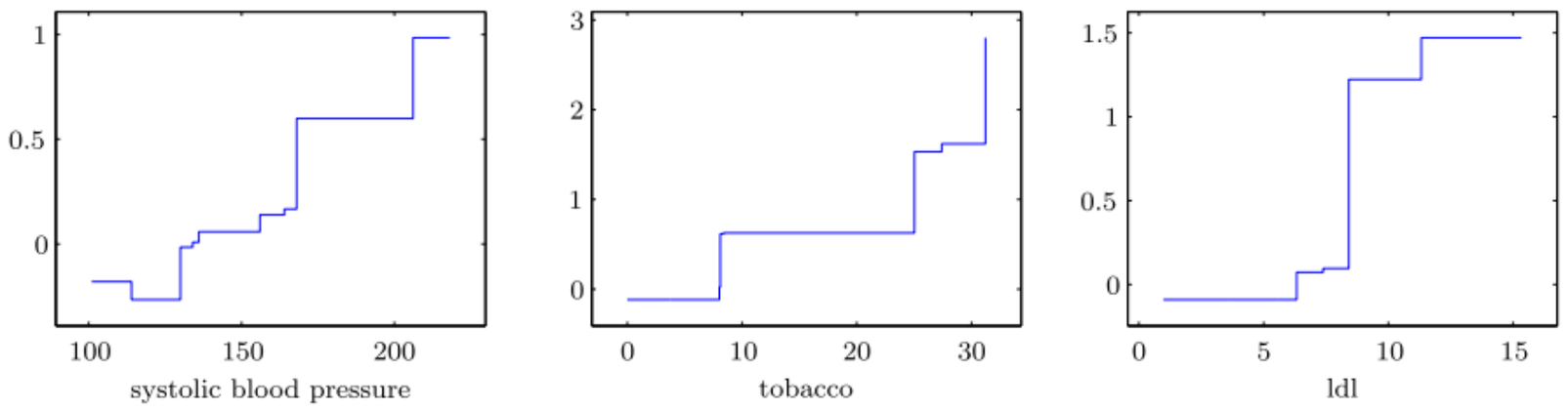
- Classify whether someone has heart disease

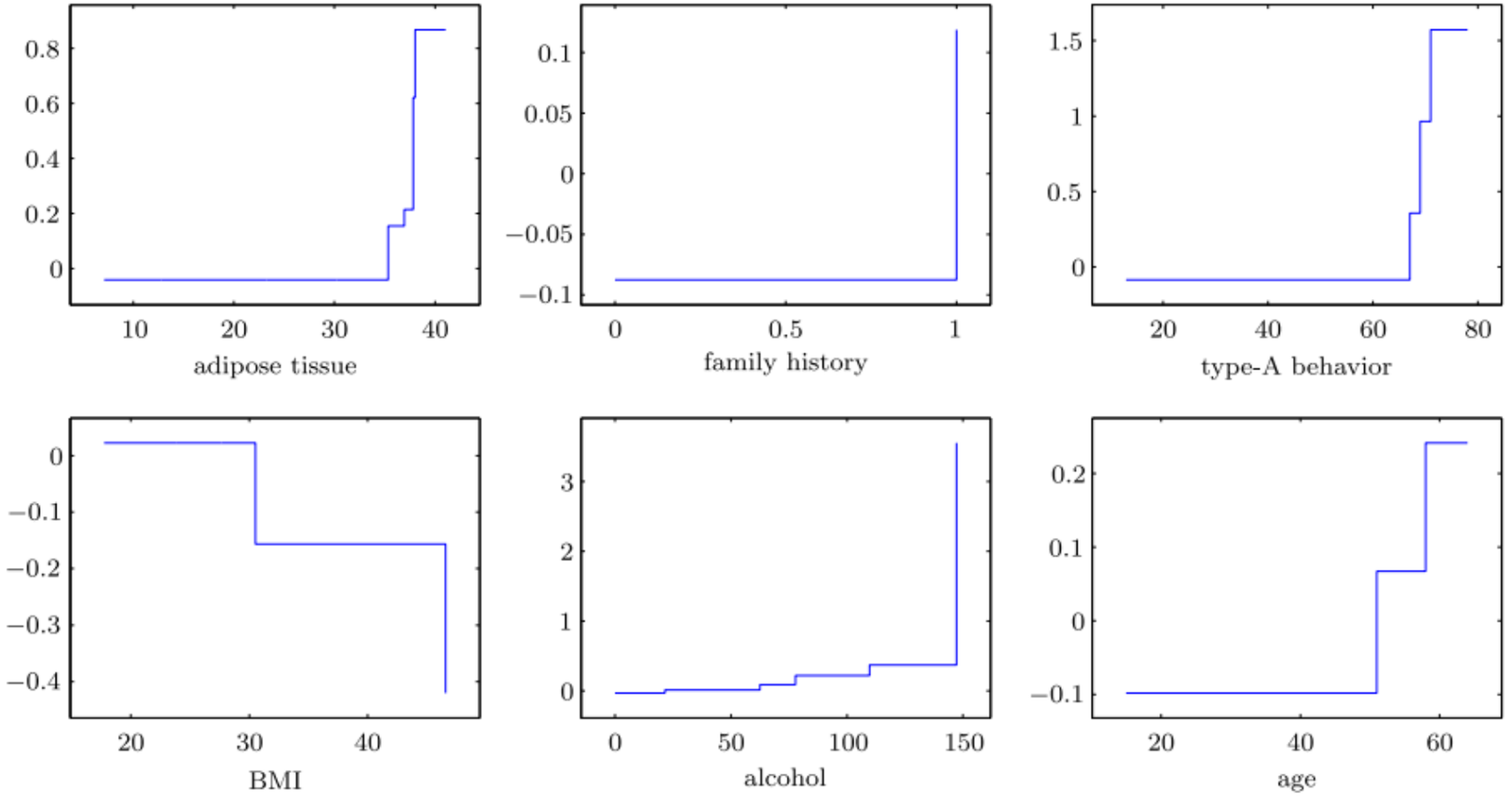
- 9 predictors including systolic blood pressure, cumulative tobacco consumption, family history, body-mass index, alcohol consumption and age

- Fit a monotone increasing piecewise constant function to each predictor

$$f_i(x; p_i) = \sum_j (p_i)_j I_{\{x \geq (x_i)_j\}}(x)$$

Example: Coronary Heart Disease Functions





Summary

- Generalized additive model can be computationally intensive task with a large number of features

- ADMM can be used to obtain a distributed approach for fitting such models

- Model can be used to promote certain properties in the fitted model such as sparsity and interpretability