ST790-003, Homework 3 (Updated Feb 17 @ 11PM)

Due Wednesday, Feb 25, 2015 @ 11:59PM

Convex or Not?

Determine whether each of the following loss functions in statistics is convex or not. If it is convex, give a rigorous proof. If it is not convex, give a counter example. A concrete numerical example is perfect as a counter example. This is a *solo* homework. Discussion with fellow students is allowed but you have to write your code and report independently.

1. (Least squares, general least squares, and nonlinear least squares) Given data $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, least squares criterion is

$$\ell(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2.$$

General least squares criterion is

$$\ell(\boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}),$$

where $\boldsymbol{\Omega}$ is a fixed $n \times n$ positive definite matrix. Nonlinear least squares criterion is

$$\ell(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - f(\boldsymbol{x}_i, \boldsymbol{\beta}))^2.$$

2. (ℓ_p regression) For a fixed $p \in [0, \infty]$, ℓ_p regression finds the regression coefficients $\boldsymbol{\beta}$ that minimizes the ℓ_p norm of residual vector

$$\ell_p(\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_p.$$

When $p = 0$, it is the best subset regression. When $p = 1$, it is the ℓ_1 regression. When $p = 2$, it is the least squares problem. When $p = \infty$, it is the ℓ_∞ regression.

3. (Worst k error regression) Define absolute residuals $r_i(\boldsymbol{\beta}) = |y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|$. If our primary interest is to reduce the k largest approximation errors, then we seek regression coefficients $\boldsymbol{\beta} \in \mathbb{R}^p$ that minimize

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{k} r_{(i)}(\boldsymbol{\beta}),$$

where $r_{(1)} \geq r_{(2)} \geq \cdots \geq r_{(n)}$ are order statistics of absolute residuals. When $k = 1$, it reduces to ℓ_∞ regression. When $k = n$, it reduces to ℓ_1 regression.

4. (Quantile regression) It is well known that the sample median of $y_1, \ldots, y_n$ is the minimizer of the sum of absolute deviations $\sum_{i=1}^n |y_i - \xi|$. Therefore ℓ_1 regression is called the median regression. Likewise, the τ-th sample quantile, $\tau \in (0, 1)$, is the minimizer of $\sum_{i=1}^n \rho_\tau(y_i - \xi)$,

where $\rho_\tau(z) = z(\tau - I_{\{z<0\}})$. This leads to the quantile regression problem that minimizes the loss function

$$\ell_\tau(\boldsymbol{\beta}) = \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}).$$

The $\tau = 1/2$ case recovers the ℓ_1 regression.

Plot the function $\rho_\tau(z)$ for a couple of τ values and determine whether the loss function $\ell_\tau(\boldsymbol{\beta})$ is convex or not.

5. (Variance component model) Consider the multivariate normal model $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Omega})$, where

$$\boldsymbol{\Omega} = \sum_{i=1}^m \sigma_i^2 \boldsymbol{V}_i$$

and $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_m$ are m fixed positive semidefinite matrices. Parameters are mean effects $\boldsymbol{\beta} \in \mathbb{R}^p$ and non-negative variance components $\sigma_1^2, \ldots, \sigma_m^2$. The negative log-likelihood function is

$$\ell(\boldsymbol{\beta}, \sigma_1^2, \ldots, \sigma_m^2) = \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln \det \boldsymbol{\Omega} + \frac{1}{2} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

6. (Linear mixed model) Consider the model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta} \in \mathbb{R}^p$ are fixed effects, $\boldsymbol{\gamma} \sim N(\boldsymbol{0}, \boldsymbol{R})$ with $\boldsymbol{R}$ unknown are random effects, and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma_0^2 \boldsymbol{I})$ are noises independent of $\boldsymbol{\gamma}$. Write down the negative log-likelihood and determine wether it is convex or not.

7. (Gaussian mixture model) Gaussian mixture model assumes that a data point $\boldsymbol{y} \in \mathbf{R}^d$ comes from multivariate Gaussian distribution $N(\boldsymbol{\mu}_j, \boldsymbol{\Omega}_j)$ with probability π_j for $j = 1, \ldots, k$. Therefore the parameters of interested are k component probabilities $(\pi_1, \ldots, \pi_k)$, k component means $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)$, and k component covariance matrices $(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_k)$. Write down the negative log-likelihood and determine wether it is convex or not.

8. (Logistic regression and GLM) The negative log-likelihood of the logistic regression model is

$$\ell(\boldsymbol{\beta}) = -\sum_{i=1}^n \left[y_i \boldsymbol{x}_i^T \boldsymbol{\beta} - \ln\left(1 + e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}\right) \right].$$

How about the negative log-likelihood of any generalized linear model (GLM) with canonical link?

9. (Gaussian covariance estimation) Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^p$ are iid multivariate normal $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. Write down the negative log-likelihood function $\ell(\boldsymbol{\Sigma})$ and determine whether it is convex or not.

10. (Gaussian precision matrix estimation) Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^p$ are iid multivariate normal $N(\boldsymbol{0}, \boldsymbol{\Omega}^{-1})$. Write down the negative log-likelihood function $\ell(\boldsymbol{\Omega})$ and determine whether it is convex or not.