Lec6-mwf

Monday, January 19, 2009 5:41 PM

HW2 is assigned

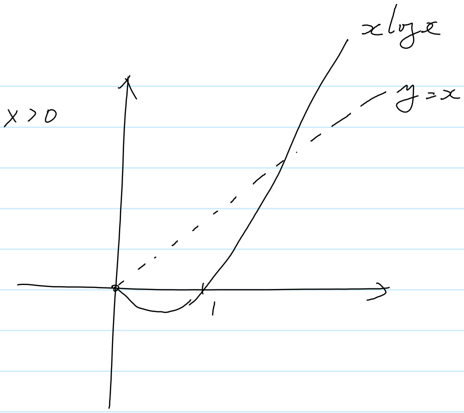
More examples

Monday, January 12, 2009 4:17 PM

(3) entropy $x \log x$, $x>0$

$f'(x) = \log x + 1$

$f''(x) = \frac{1}{x} + 1$



The source generates symbols according a distribution p.

$$p\{\text{symbox } X_i\} = p_i \quad , i=1,\cdots,N$$

Entropy

$$H = -\sum_{i=1}^{N} p_i \log p_i$$

measures the "uncertainty" of the source (i.e, the amount of information generated by the source.)

- KL divergence (Kullback-Leibler)
  - Two distributions: (p_i) & $[q_i]$

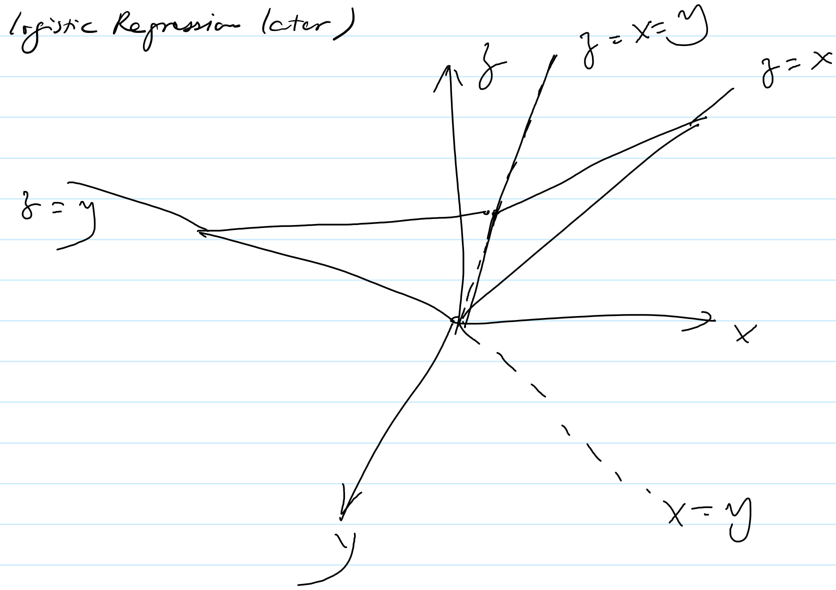
    $$D_{KL}(P\|Q) = \sum p_i \log \frac{p_i}{q_i}$$

  - If (q_i) is fixed, then this is convex is (p_i).

  But in fact, this is convex in both (p_i) & (q_i)
  - Will need the "perspective" of a function later
  - Useful for measuring the "distance" between distributions. (See logistic Regression later)

(4) max:

$\max_i x_i$ is convex



- non-differentiable!
- need the "max. of convex functions"

(5) log-sum-exp

$\log(e^{x_1} + \cdots + e^{x_N})$ is convex

Proof: Restrict to a line and use second-order conditions.

$$g(\lambda) = f(\vec{x} + \lambda \vec{y}) = \log\left[e^{x_1+\lambda y_1} + \cdots + e^{x_N+\lambda y_N}\right]$$

$$\frac{df}{d\lambda} = \frac{y_1 e^{x_1+\lambda y_1} + \cdots + y_N e^{x_N+\lambda y_N}}{e^{x_1+\lambda y_1} + \cdots + e^{x_N+\lambda y_N}}$$

$$\frac{d^2 f}{d\lambda^2} = \frac{1}{(e^{x_1+\lambda y_1} + \cdots + e^{x_N+\lambda y_N})^2} \cdot$$

$$\Big\{ (y_1^2 e^{x_1+\lambda y_1} + \cdots + y_N^2 e^{x_N+\lambda y_N}) \cdot (e^{x_1+\lambda y_1} + \cdots + e^{x_N+\lambda y_N}) - (y_1 e^{x_1+\lambda y_1} + \cdots + y_N e^{x_N+\lambda y_N})^2 \Big\}$$

Use Cauchy-Schwartz Inequality.

$$\left(\sum a_n^2\right)\left(\sum b_n^2\right) \geq \left(\sum a_n b_n\right)^2$$

Use $a_n^2 = y_n^2 e^{x_n+\lambda y_n}$, $b_n^2 = e^{x_n+\lambda y_n}$

$n = 1, 2, \cdots, N.$

Example:

Skip

ⓐ log-moment-generating function

$$\log E[e^{sX}] = \log\left(\sum_i P_i e^{sx_i}\right) \quad \text{convex in } s.$$

$$P\{X = x_i\} = P_i$$

Skip for now.

ⓑ high-SNR approximation of Shannon capacity

$$C_i = W \log\left(1 + \frac{P_i}{\sum_{j \neq i} P_j + N}\right)$$

interference noise

At high-SNR, $\log(1+x) \approx \log x$

$$C_i \approx W \cdot \log \frac{P_i}{\sum_{j\neq i} P_j + N}$$

$$= W \cdot \log P_i - \log\left(\sum_{j\neq i} P_j + N\right)$$

This is neither concave not convex!

Let $p_i = e^{x_i}$

$$C = W \cdot x_i - \log\left(\sum_{j \neq i} e^{x_i} + N\right)$$

concave in $\vec{x}$!

(c) low-SINR approximation of Shannon Capacity

Skip

When x is small, $\log(1+x) \simeq x$

$$C_i \simeq W \cdot \frac{p_i}{\sum_{j \neq i} p_j + N}$$

- linear in p_i
- convex in p_j, $j \neq i$
- NOT convex in $\vec{p} = (p_i)_{i=1,\dots,N}$

Let $y_i = \log C_i$,

$$y_i = \log W + \log \frac{p_i}{\sum_{j \neq i} p_i + N}$$

Let $p_i = e^{x_i}$

$$y_i = \log W + x_i - \log\left(\sum_{j \neq i} e^{x_j} + N\right)$$

which is concave in $\vec{x}$.

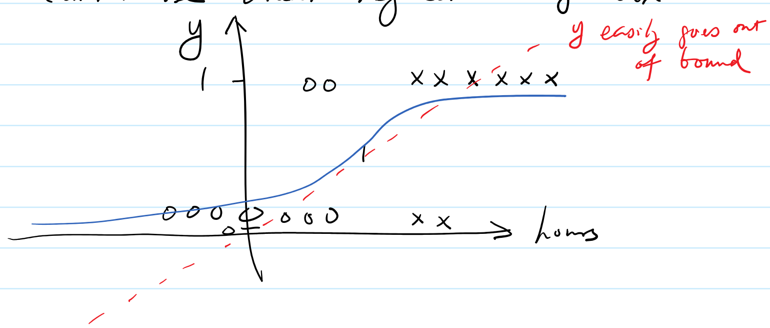
(d) Logistic Regression.

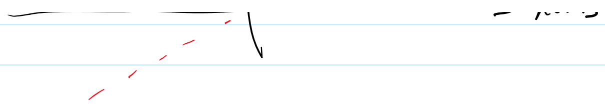
- Suppose you have training data (x_i, y_i)

x_i: explanatory variable, e.g. hours of study

y_i: categorical variable (i.e., label), e.g. pass/fail (1/0)

- We want to learn how to predict y from x
  - cannot use linear regression $y = ax$





- Instead, use something like the blue curve.

$$y = \frac{e^{ax+b}}{1+e^{ax+b}}$$

approach 0 if $ax \to -\infty$
approach 1 if $ax \to +\infty$

- This can be thought of as the conditional distribution of y (1/0) given x.
  - $\overset{1}{\underset{0}{\gtrless}} 0.5$

- We need to use a distance to measure how good the match is.

  - Cannot use L2 - distance

$$\left(y - \frac{e^{ax+b}}{1+e^{ax+b}} \right)^2$$

is not a convex function!

- Use KL divergence
  - P : empirical observation of the conditional distribution of y given x
  - Q : our estimate above

  - $$D_{KL}(P \,\|\, Q) = \sum P_k \log \frac{P_k}{Q_k}$$ convex in (P, Q)

$$= \underbrace{\sum_k P_k \log P_k}_{\text{indep of } a} - \sum_k P_k \log Q_k$$

- We thus maximize $\sum_k P_k \log Q_k$

  - k denote a possible outcome.

  - Suppose the sample is (x_i, y_i)

  - Given x_i, our P or Q places probability on two outcomes 1 or 0.

  - For P, it is
    $(1, 0)$ if $y_i = 1$
    $(0, 1)$ if $y_i = 0$

- For P, it is

$$(1, 0) \quad \text{if } y_i = 1$$
$$(0, 1) \quad \text{if } y_i = 0$$

- Together

$$\left(1_{\{y_i=1\}},\ 1_{\{y_i=0\}} \right)$$

- For Q, it is

$$\left(\frac{e^{ax_i+b}}{1+e^{ax_i+b}},\ \frac{1}{1+e^{ax_i+b}} \right)$$

- The KL-divergence is then

$$1_{\{y_i=1\}} \cdot \log \frac{e^{ax_i+b}}{1+e^{ax_i+b}}$$
$$+ 1_{\{y_i=0\}} \log \frac{1}{1+e^{ax_i+b}}$$

- This is the logistic regression:

$$\max \sum_i \left\{ 1_{\{y_i=1\}} \log \frac{e^{ax_i+b}}{1+e^{ax_i+b}} + 1_{\{y_i=0\}} \log \frac{1}{1+e^{ax_i+b}} \right\}$$

- We can verify that it is concave in a.

⑥ Geometric mean

$$\left(\prod_{i=1}^{N} x_i \right)^{1/N} \quad \text{is concave}$$

Proof: Similar to ⑤. See Boyd P74.

⑦ Quadratic-over-Linear

$$\frac{x^2}{y} \quad \text{is convex}$$

Proof: $\nabla^2 = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} \geq 0.$

Ⓠ Why is it positive-semidefinite?

- Can also use "perspective mapping" (later)

More operations

Wednesday, January 14, 2009 10:20 AM

Earlier we have discussed two operations that preserve convexity.

- Non-negative weighted sums
- Affine mapping of the arguments.

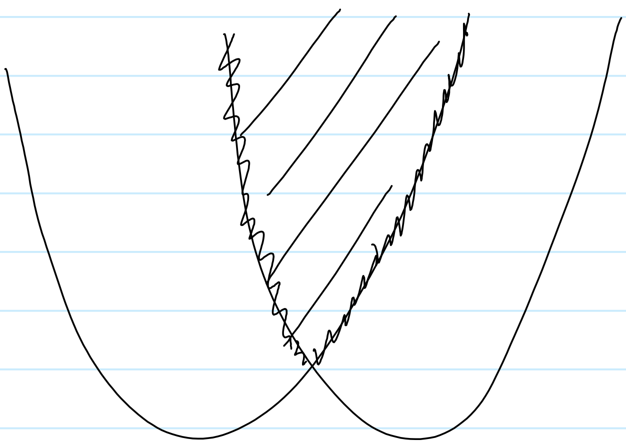
③ Pointwise maximum & supremum

If f_1, f_2 are convex, define

$$f(x) \triangleq \max\{f_1(x), f_2(x)\}$$

Then $f(x)$ is convex

Proof: The epigraph of $f(x)$ is the intersection of the epigraphs of f_1 & f_2.



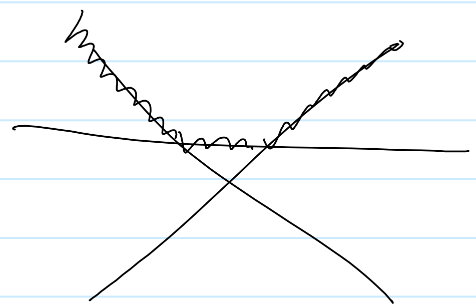
The supremum can be taken over an infinite set.

If $f(x,y)$ is convex in x for each y
then

$$g(x) = \sup_{y \in A} f(x, y) \text{ is convex in } x.$$

Similarly, pointwise minimum & infimum of concave functions is a concave function.

Examples:



(a) $\max_i X_i$ is convex

$\max_i a_i^T x_i + b_i$ is convex

$\min_i X_i$ is concave

$\min_i a_i^T x_i + b_i$ is concave

skip (b) Sum of the largest r components

For $x \in R^n$, let

$$x_{[1]} \geq x_{[2]} \geq \cdots x_{[n]}$$

where $[1], \cdots [n]$ is a permutation of 1 to n.

Let

$$f(x) = \sum_{i=1}^{r} x_{[i]}$$

Then $f(x)$ is convex

$$f(x) = \max \left(x_{i_1} + \cdots + x_{i_r} \,\middle|\, 1 \leq i_1 < i_2 < \cdots < i_r \leq n \right)$$

(c) Distance to the furthest point.

Let x be R^n, $C \subseteq R^n$

Let $\|x-y\|$ denote the distance btw x and y.

Let $f(x) = \sup_{y \in C} \|x-y\|$

Then $f(x)$ is convex

(d) Infimum of convex functions?

$\inf_{y \in C} \|x-y\|$?

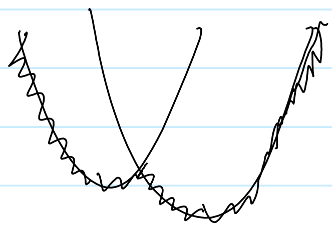
(10)

Minimizations

Wednesday, January 14, 2009 10:34 AM

(4) Minimization of a jointly convex function over a convex set.

In general, minimization of convex function is NOT a convex function.



However, there is an important exception.

If f is convex in (x,y) jointly. and C is a convex non-empty set, then

$$g(x) \triangleq \inf_{y \in C} f(x,y). \quad \text{is convex.}$$

Difference with maximization

- f must be jointly convex
- Must take the minimum over a convex set (e.g. can't just take two functions)

Proof: Assume that the infimum is always attained. (What to do if this assumption does not hold?)

Take $g(x_1) \triangleq \inf_{y \in C} f(x_1, y) = f(x_1, y_1)$

$g(x_2) \triangleq \inf_{y \in C} f(x_2, y) = f(x_2, y_2)$

We want to show that

$$g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta) g(x_2)$$

$$\overset{\triangle}{=}$$

$$\inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2)$$

It suffices to find a $y \in C$ such that

$$f(\theta x_1 + (1-\theta)x_2, y) \leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2)$$

Taking $y = \theta y_1 + (1-\theta) y_2$, we have

$y \in C$ and the result follows from the convexity of f.

Examples:

ⓐ Distance to a set

$$\text{dist}(x, C) \triangleq \inf_{y \in C} \|x - y\|$$

If C is a convex set, then $\text{dist}(x, C)$ is convex.

(b) $g(x) = \inf f_0(y)$

sub $f_1(y) \le x$

If f_0 & f_1 convex, then g is convex

Proof:

Define $f(x,y) = \begin{cases} f_0(y) & \text{if } f_1(y) \le x \\ +\infty & \text{otherwise} \end{cases}$

- This is the "extended" function that we will see again in composition

Then $f(x,y)$ is convex in (x,y) (proved by yourself)

& $g(x) = \inf_{y \in R} f(x,y)$

- Useful when we study strong duality.

25