Applications (I)

Lijun Zhang
zlj@nju.edu.cn
http://cs.nju.edu.cn/zlj

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

Basic Norm Approximation

Norm Approximation Problem

$$\min \quad \|Ax - b\|$$

- $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$ are problem data
- $x \in \mathbf{R}^n$ is the variable
- $\|\cdot\|$ is a norm on $\mathbf{R}^n$
- Approximation solution of $Ax \approx b$, in $\|\cdot\|$

Residual

$$r = Ax - b$$

A Convex Problem

- $b \in \mathcal{R}(A)$, the optimal value is 0
- $b \notin \mathcal{R}(A)$, more interesting

Basic Norm Approximation

Approximation Interpretation

$$Ax = x_1 a_1 + \cdots + x_n a_n$$

- $a_1, \dots, a_n \in \mathbf{R}^m$ are the columns of A
- Approximate the vector b by a linear combination

- Regression problem
  - ✓ $a_1, \dots, a_n$ are regressors
  - ✓ $x_1 a_1 + \cdots + x_n a_n$ is the regression of b

Basic Norm Approximation

Estimation Interpretation

- Consider a linear measurement model

$$y = Ax + v$$

- $y \in \mathbf{R}^m$ is a vector measurement
- $x \in \mathbf{R}^n$ is a vector of parameters to be estimated
- $v \in \mathbf{R}^m$ is some measurement error that is unknown, but presumed to be small
- Assume smaller values of v are more plausible

$$\hat{x} = \mathrm{argmin}_z \|Az - y\|$$

Basic Norm Approximation

Geometric Interpretation

- Consider the subspace $\mathcal{A} = \mathcal{R}(A) \subseteq \mathbf{R}^m$, and a point $b \in \mathbf{R}^m$
- A projection of the point b onto the subspace $\mathcal{A}$, in the norm $\|\cdot\|$

$$\begin{array}{ll} \min & \|u - b\| \\ \text{s.t.} & u \in \mathcal{A} \end{array}$$

- Parametrize an arbitrary element of $\mathcal{R}(A)$ as $u = Ax$, we see that norm approximation is equivalent to projection

Basic Norm Approximation

- Weighted Norm Approximation Problems

$$\min \quad \|W(Ax - b)\|$$

  - $W \in \mathbf{R}^{m \times m}$ is called the weighting matrix
  - A norm approximation problem with norm $\|\cdot\|$, and data $\tilde{A} = WA, \tilde{b} = Wb$
  - A norm approximation problem with data A and b, and the W-weighted norm

$$\|z\|_W = \|Wz\|$$

Basic Norm Approximation

□ Least-Squares Approximation

$$\min \quad \|Ax - b\|_2^2 = r_1^2 + r_2^2 + \cdots + r_m^2$$

- The minimization of a convex quadratic function

$$f(x) = x^\top A^\top A x - 2b^\top A x + b^\top b$$

- A point x minimizes f if and only if

$$\nabla f(x) = 2A^\top A x - 2A^\top b = 0$$

- Normal equations

$$A^\top A x = A^\top b$$

Basic Norm Approximation

- Chebyshev or Minimax Approximation

$$\min \quad \|Ax - b\|_\infty = \max\{|r_1|, \dots, |r_m|\}$$

  - Be cast as an LP

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & -t\mathbf{1} \preccurlyeq Ax - b \preccurlyeq t\mathbf{1} \end{aligned}$$

  with variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$

- Sum of Absolute Residuals Approximation

$$\min \quad \|Ax - b\|_1 = |r_1| + \cdots + |r_m|$$

  - Be cast as an LP

$$\begin{aligned} \min \quad & \mathbf{1}^\top t \\ \text{s.t.} \quad & -t \preccurlyeq Ax - b \preccurlyeq t \end{aligned}$$

  with variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}^m$

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

l_p-norm Approximation

- l_p-norm approximation, for $1 \leq p \leq \infty$

$$(|r_1|^p + \cdots + |r_m|^p)^{1/p}$$

- The equivalent problem with objective

$$|r_1|^p + \cdots + |r_m|^p$$

  - A separable and symmetric function of the residuals
  - Objective depends only on the amplitude distribution of the residuals

Penalty Function Approximation

The Problem

$$\begin{array}{ll} \min & \phi(r_1) + \cdots + \phi(r_m) \\ \text{s.t.} & r = Ax - b \end{array}$$

- $\phi: \mathbf{R} \to \mathbf{R}$ is called the penalty function
- ϕ is convex
- ϕ is symmetric, nonnegative, and satisfies $\varphi(0) = 0$

- A penalty function assesses a cost or penalty for each component of residual

Example

- ℓ_p-norm Approximation

$$\phi(u) = |u|^p$$

  - Quadratic penalty: $\phi(u) = u^2$
  - Absolute value penalty: $\phi(u) = |u|$

- Deadzone-linear Penalty Function

$$\phi(u) = \begin{cases} 0 & |u| \le a \\ |u| - a & |u| > a \end{cases}$$

- The Log Barrier Penalty Function

$$\phi(u) = \begin{cases} -a^2 \log\left(1 - (u/a)^2\right) & |u| < a \\ \infty & |u| \ge a \end{cases}$$

Example

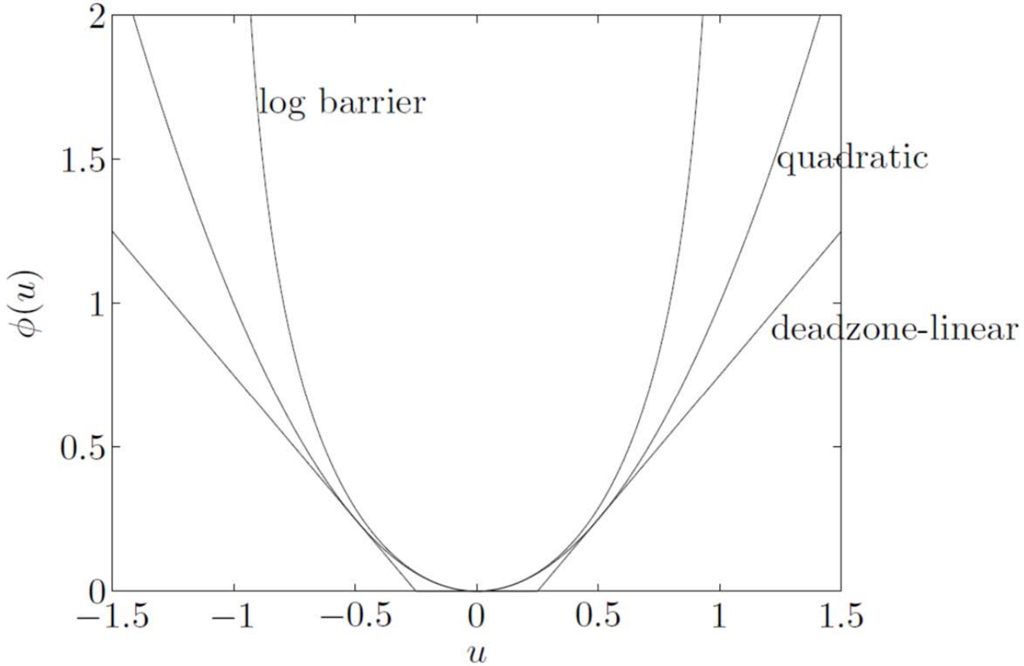


Figure 6.1 Some common penalty functions: the quadratic penalty function $\phi(u) = u^2$, the deadzone-linear penalty function with deadzone width $a = 1/4$, and the log barrier penalty function with limit $a = 1$.

- Log barrier penalty function assesses an infinite penalty for residuals larger than a
- Log barrier function is very close to the quadratic penalty for $|u/a| \leq 0.25$

Discussions

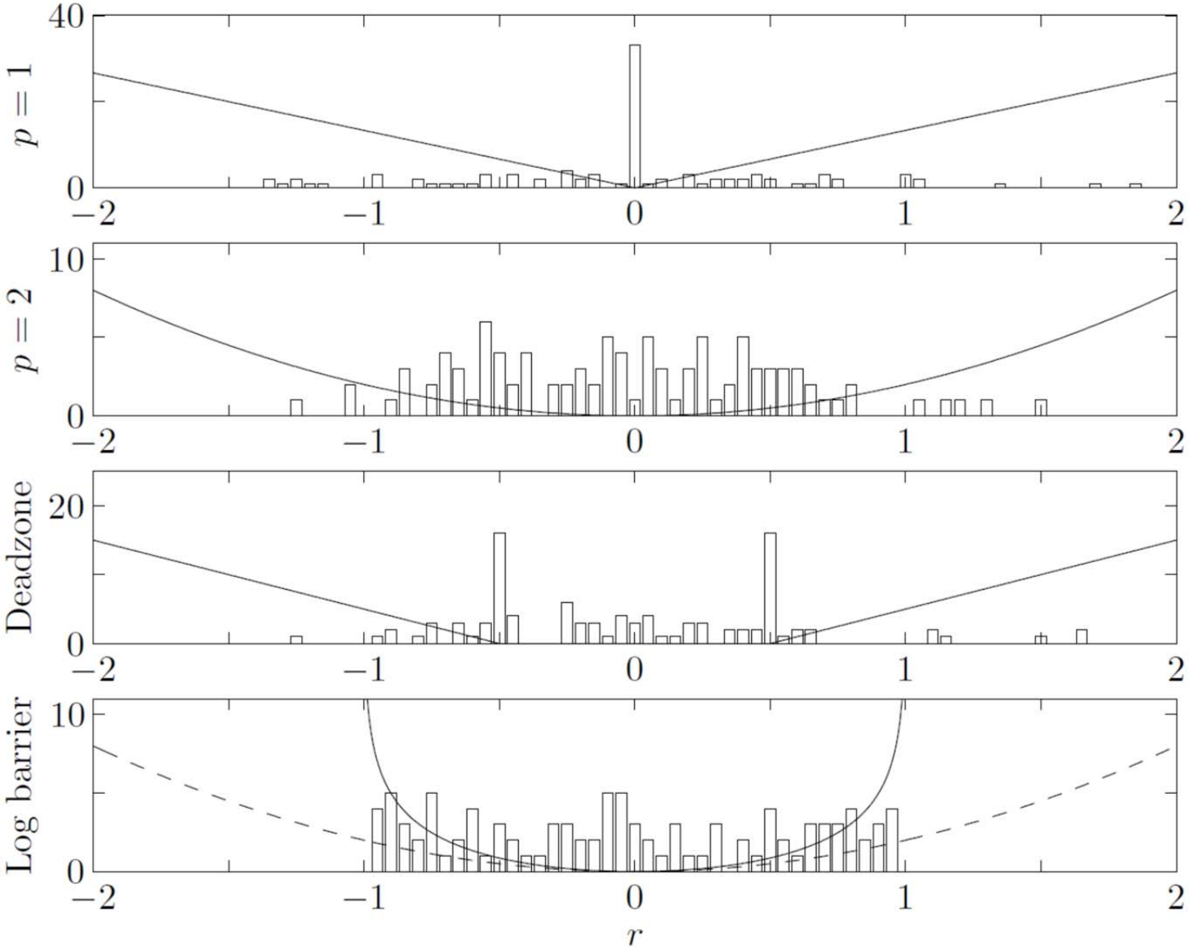
- Roughly speaking, $\varphi(u)$ is a measure of our dislike of a residual of value u
- If φ is very small for small u, it means we care very little if residuals have these values
- If $\varphi(u)$ grows rapidly as u becomes large, it means we have a strong dislike for large residuals
- If φ becomes infinite outside some interval, it means that residuals outside the interval are unacceptable

Discussions

- $\phi_1(u) = |u|$、$\phi_2(u) = u^2$
  - For small u we have $\phi_1(u) \gg \phi_2(u)$, so ℓ_1-norm approximation puts relatively larger emphasis on small residuals
  - The optimal residual for the ℓ_1-norm approximation problem will tend to have more zero and very small residuals
  - For large u we have $\phi_2(u) \gg \phi_1(u)$, so ℓ_1-norm approximation puts less weight on large residuals
  - The ℓ_2-norm solution will tend to have relatively fewer large residuals

Example

- $A \in \mathbf{R}^{100\times 30}$, b$\in \mathbf{R}^{100}$



Observations of Penalty Functions

- The ℓ_1-norm penalty puts the most weight on small residuals and the least weight on large residuals.
- The ℓ_2-norm penalty puts very small weight on small residuals, but strong weight on large residuals.
- The deadzone-linear penalty function puts no weight on residuals smaller than 0.5, and relatively little weight on large residuals.
- The log barrier penalty puts weight very much like the ℓ_2-norm penalty for small residuals, but puts very strong weight on residuals larger than around 0.8, and infinite weight on residuals larger than 1.

Observations of Amplitude Distributions

- For the ℓ_1-optimal solution, many residuals are either zero or very small. The ℓ_1-optimal solution also has relatively more large residuals.
- The ℓ_2-norm approximation has many modest residuals, and relatively few larger ones.
- For the deadzone-linear penalty, we see that many residuals have the value ± 0.5, right at the edge of the 'free' zone, for which no penalty is assessed.
- For the log barrier penalty, we see that no residuals have a magnitude larger than 1, but otherwise the residual distribution is similar to the residual distribution for ℓ_2-norm approximation.

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

Approximation with Constraints

- Add Constraints to

$$\min \quad \|Ax - b\|$$

  - Rule out certain unacceptable approximations of the vector b
  - Ensure that the approximator Ax satisfies certain properties
  - Prior knowledge of the vector x to be estimated
  - Prior knowledge of the estimation error v
  - Determine the projection of a point b on a set more complicated than a subspace

Approximation with Constraints

- Nonnegativity Constraints on Variables

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & x \succcurlyeq 0 \end{aligned}$$

  - Estimate a vector x of parameters known to be nonnegative
  - Determine the projection of a vector b onto the cone generated by the columns of A
  - Approximate b using a nonnegative linear combination of the columns of A

Approximation with Constraints

Variable Bounds

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & l \leqslant x \leqslant u \end{aligned}$$

- Prior knowledge of intervals in which each variable lies

- Determine the projection of a vector b onto the image of a box under the linear mapping induced by A

Approximation with Constraints

Probability Distribution

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & x \succcurlyeq 0, \mathbf{1}^{\top} x = 1 \end{aligned}$$

- Estimation of proportions or relative frequencies
- Approximate b by a convex combination of the columns of A

Norm Ball Constraint

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & \|x - x_0\| \leq d \end{aligned}$$

- x_0 is prior guess of what the parameter x is, and d is the maximum plausible deviation

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

Least-norm Problems

Basic least-norm Problem

$$\begin{aligned} \min \quad & \|x\| \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

- $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$
- $x \in \mathbf{R}^n$, $\|\cdot\|$ is a norm on $\mathbf{R}^n$
- The solution is called a least-norm solution of $Ax = b$.
- A convex optimization problem
- Interesting when $m \leq n$

Least-norm Problems

- Reformulation as Norm Approximation Problem
  - Let x_0 be any solution of $Ax = b$
  - Let $Z \in \mathbf{R}^{n \times k}$ be a matrix whose columns are a basis for the nullspace of A.

$$\{x | Ax = b\} = \{x_0 + Zu | u \in \mathbf{R}^k\}$$

  - The least-norm problem can be expressed as

$$\min \quad \|x_0 + Zu\|$$

Least-norm Problems

Estimation interpretation

- We have $m < n$ perfect linear measurement, given by $Ax = b$
- Our measurements do not completely determine x

- Suppose our prior information, is that x is more likely to be small than large.
- Choose the parameter vector x which is smallest among all parameter vectors that are consistent with the measurements

Least-norm Problems

- Geometric interpretation
  - The feasible set $\{x|Ax = b\}$ is affine
  - The objective is the distance between x and the point 0

  - Find the point in the affine set with minimum distance to 0
  - Determine the projection of the point 0 on the affine set $\{x|Ax = b\}$

Least-norm Problems

- Least-squares Solution of Linear Equations

$$\begin{aligned} \min \quad & \|x\|_2^2 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

  - The optimality conditions

$$2x^* + A^\top v^* = 0 \quad Ax^* = b$$

    - ✓ v is the dual variable
  - The Solution

$$x^* = -\frac{1}{2}A^\top v^* \Rightarrow -\frac{1}{2}AA^\top v^* = b$$

$$\Rightarrow \quad v^* = -2(AA^\top)^{-1}b, x^* = A^\top(AA^\top)^{-1}b$$

Least-norm Problems

Least-penalty Problems

$$\begin{array}{ll} \min & \phi(x_1) + \cdots + \phi(x_n) \\ \text{s.t.} & Ax = b \end{array}$$

- $\phi: \mathbf{R} \to \mathbf{R}$ is convex, nonnegative and satisfies $\phi(0) = 0$
- The penalty function value $\phi(u)$ quantifies our dislike of a component of x having value u
- Find x that has least total penalty, subject to the constraint $Ax = b$

Least-norm Problems

□ Sparse Solutions via Least ℓ_1-norm

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

- Tend to produce a solution x with a large number of components equal to 0
- Tend to produce sparse solutions of $Ax = b$, often with m nonzero components

Least-norm Problems

- Sparse Solutions via Least ℓ_1-norm

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

- Find solutions of $Ax = b$ that have only m nonzero components
  - $\tilde{A}$ is a submatrix of A
  - $\tilde{x}$ and subvector of x
  - Solve $\tilde{A}\tilde{x} = b$
    - ✓ If there is a solution, we are done
  - Complexity: $n!/(m!\,(n-m)!)$

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

Bi-criterion Formulation

- A (convex) Vector Optimization Problem with Two Objectives

$$\min(\text{w.r.t.}\, \mathbf{R}_+^2) \quad (\|Ax - b\|, \|x\|)$$

  - Find a vector x that is small
  - Make the residual $Ax - b$ small
  - Optimal trade-off between the two objectives
    - ✓ The minimum value of $\|x\|$ is 0 and the residual norm is $\|b\|$
    - ✓ Let C denote the set of minimizers of $\|Ax - b\|$, and then any minimum norm point in C is Pareto optimal

Regularization

- Weighted Sum of the Objectives

$$\min \quad \|Ax - b\| + \gamma \|x\|$$

  - $\gamma > 0$ is a problem parameter
  - A common scalarization method used to solve the bi-criterion problem
  - As γ varies over $(0, \infty)$, the solution traces out the optimal trade-off curve

- Weighted Sum of Squared Norms

$$\min \quad \|Ax - b\|^2 + \gamma \|x\|^2$$

Regularization

Tikhonov Regularization

$$\min \quad \|Ax - b\|_2^2 + \delta\|x\|_2^2 = x^\top(A^\top A + \delta I)x - 2b^\top Ax + b^\top b$$

- Analytical solution

$$x = (A^\top A + \delta I)^{-1}A^\top b$$

- Since $A^\top A + \delta I \succ 0$ for any $\delta \succ 0$, the Tikhonov regularized least-squares solution requires no rank assumptions on the matrix A

Regularization

- ℓ_1-norm Regularization

$$\min \quad \|Ax - b\|_2 + \gamma \|x\|_1$$

  - Find a sparse solution
  - The residual is measured with the Euclidean norm and the regularization is done with an ℓ_1-norm
  - By varying the parameter γ we can sweep out the optimal trade-off curve between $\|Ax - b\|_2$ and $\|x\|_1$

Example

Regressor Selection Problem

$$\begin{aligned} \min \quad & \|Ax - b\|_2 \\ \text{s.t.} \quad & \text{card}(x) \leq k \end{aligned}$$

- One straightforward approach is to check every possible sparsity pattern in x with k nonzero entries
- For a fixed sparsity pattern, we can find the optimal x by solving a least-squares problem
- Complexity: $n!/(k!\,(n-k)!)$

Example

Regressor Selection Problem

$$\begin{aligned} \min \quad & \|Ax - b\|_2 \\ \text{s.t.} \quad & \operatorname{card}(x) \leq k \end{aligned}$$

- A good heuristic approach is to solve the following problem for different γ

$$\min \quad \|Ax - b\|_2 + \gamma \|x\|_1$$

- Find the smallest value of γ that results in a solution with $\operatorname{card}(x) \leq k$
- We then fix this sparsity pattern and find the value of x that minimizes $\|Ax - b\|_2$

Example

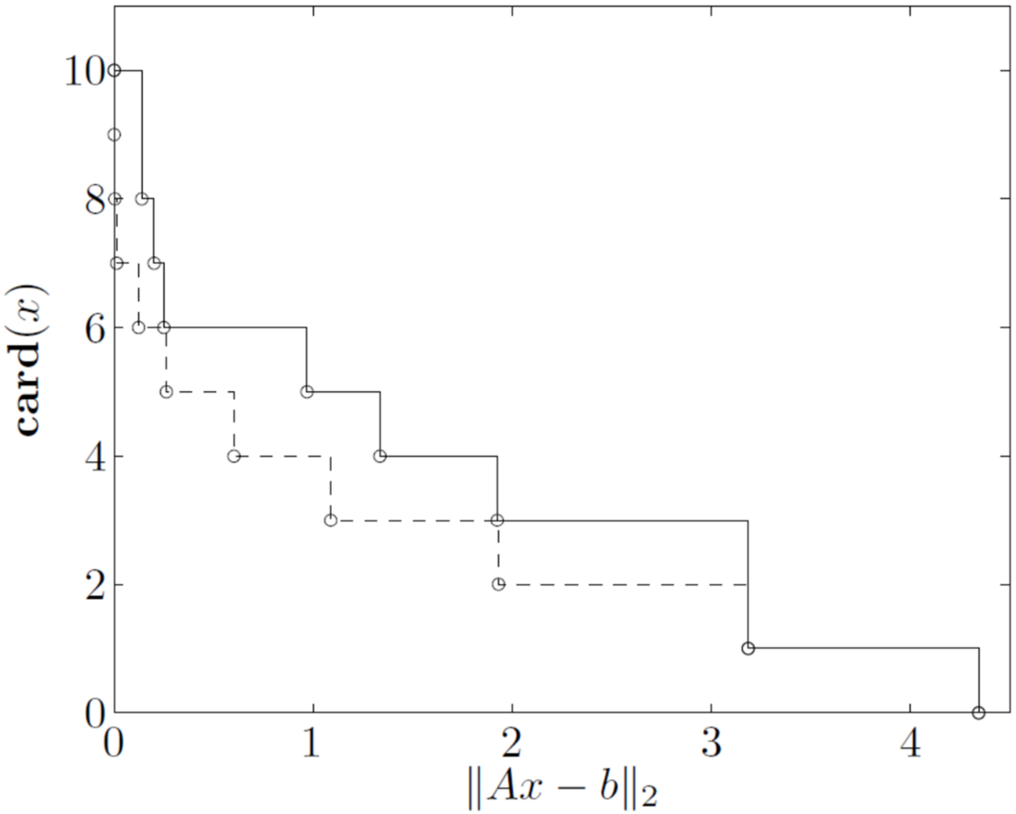


Figure 6.7 Sparse regressor selection with a matrix $A \in \mathbf{R}^{10\times 20}$. The circles on the dashed line are the Pareto optimal values for the trade-off between the residual $\|Ax - b\|_2$ and the number of nonzero elements $\mathbf{card}(x)$. The points indicated by circles on the solid line are obtained via the ℓ_1-norm regularized heuristic.

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

Classification

- Given two sets of points in $\mathbf{R}^n$

  $\{x_1, \dots, x_N\}$ and $\{y_1, \dots, y_M\}$

- Find a function $f: \mathbf{R}^n \longrightarrow \mathbf{R}$

  $f(x_i) > 0, i = 1, \dots, N, \qquad f(y_i) < 0, i = 1, \dots, M$

  - Positive on the first set and negative on the second

  - f or its 0-level set $\{x | f(x) = 0\}$, separates, classifies, or discriminates the two sets of points

Linear Discrimination

- Affine function $f(x) = a^\top x - b$

$$a^\top x_i - b > 0, i = 1, \dots, N,$$
$$a^\top y_i - b < 0, i = 1, \dots, M$$

  - A hyperplane that separates the two sets of points

- The strict inequalities are homogeneous in a and b
  - Equivalent conditions

$$a^\top x_i - b \geq 1, i = 1, \dots, N,$$
$$a^\top y_i - b \leq -1, i = 1, \dots, M$$

Example

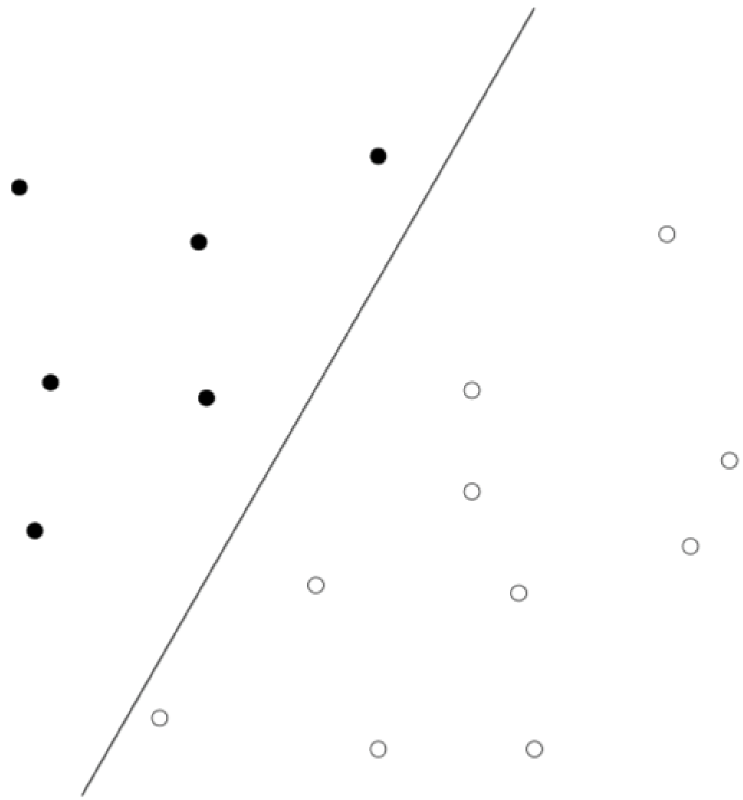


Figure 8.8 The points $x_1, \ldots, x_N$ are shown as open circles, and the points $y_1, \ldots, y_M$ are shown as filled circles. These two sets are classified by an affine function f, whose 0-level set (a line) separates them.

Robust Linear Discrimination

- Seek the function that gives the maximum possible 'gap' between x_i and y_i

$$
\begin{aligned}
\max \quad & t \\
\text{s.t.} \quad & a^\top x_i - b \geq t, i = 1, \ldots, N \\
& a^\top y_i - b \leq -t, i = 1, \ldots, M \\
& \|a\|_2 \leq 1
\end{aligned}
$$

  - a is normalized
  - The optimal value t^* is positive if and only if the two sets of points can be linearly discriminated

Example

- If $\|a\|_2 = 1$, $a^\top x_i - b$ is the Euclidean distance from the point x_i to the separating hyperplane $a^\top z = b$
- $b - a^\top y_i$ is the distance from y_i to the hyperplane

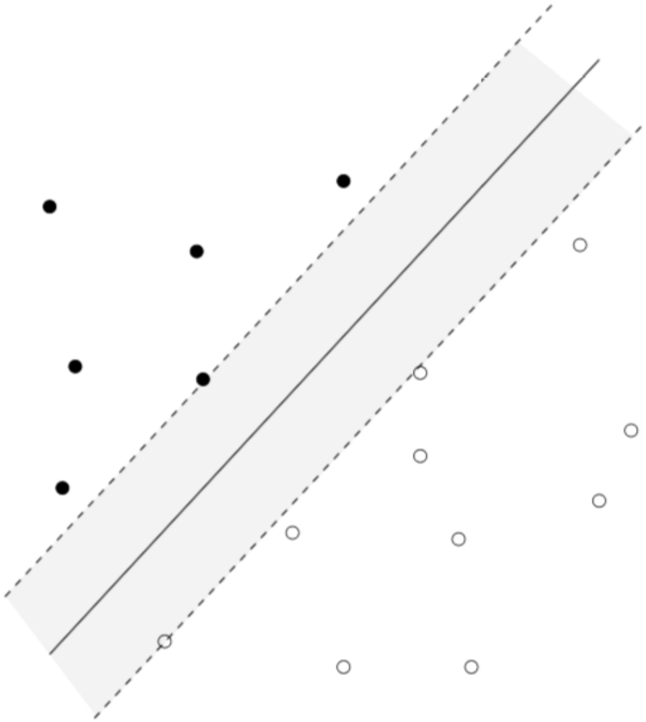


Figure 8.9 By solving the robust linear discrimination problem (8.23) we find an affine function that gives the largest gap in values between the two sets (with a normalization bound on the linear part of the function). Geometrically, we are finding the thickest slab that separates the two sets of points.

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

Support Vector Classifier

- When the two sets of points cannot be linearly separated

- One that minimizes the number of points misclassified
  - Unfortunately, this is in general a difficult combinatorial optimization problem

Support Vector Classifier

- When the two sets of points cannot be linearly separated
- Relaxation

$$a^\top x_i - b \geq 1, i = 1, \dots, N,$$
$$a^\top y_i - b \leq -1, i = 1, \dots, M$$

$$a^\top x_i - b \geq 1 - u_i, i = 1, \dots, N,$$
$$a^\top y_i - b \leq -(1 - v_i), i = 1, \dots, M$$

  - Nonnegative variables $u_1, \dots, u_N$ and $v_1, \dots, v_M$
  - When $u = v = 0$, we recover the original constraints
  - By making u and v large enough, these inequalities can always be made feasible

Support Vector Classifier

- Our goal is to find a, b and sparse nonnegative u and v that satisfy the inequalities
- We can minimize the sum of the variables u_i and v_{i}

$$
\begin{array}{ll}
\min & 1^\top u + 1^\top v \\
\mathrm{s.t.} & a^\top x_i - b \geq 1 - u_i, i = 1, \ldots, N \\
& a^\top y_i - b \leq -(1 - v_i), i = 1, \ldots, M \\
& u \succcurlyeq 0,\ v \succcurlyeq 0
\end{array}
$$

  - When $0 < u_i < 1$, x_i is classified correctly by $a^\top x - b$, but still incurs a loss u_i

Example

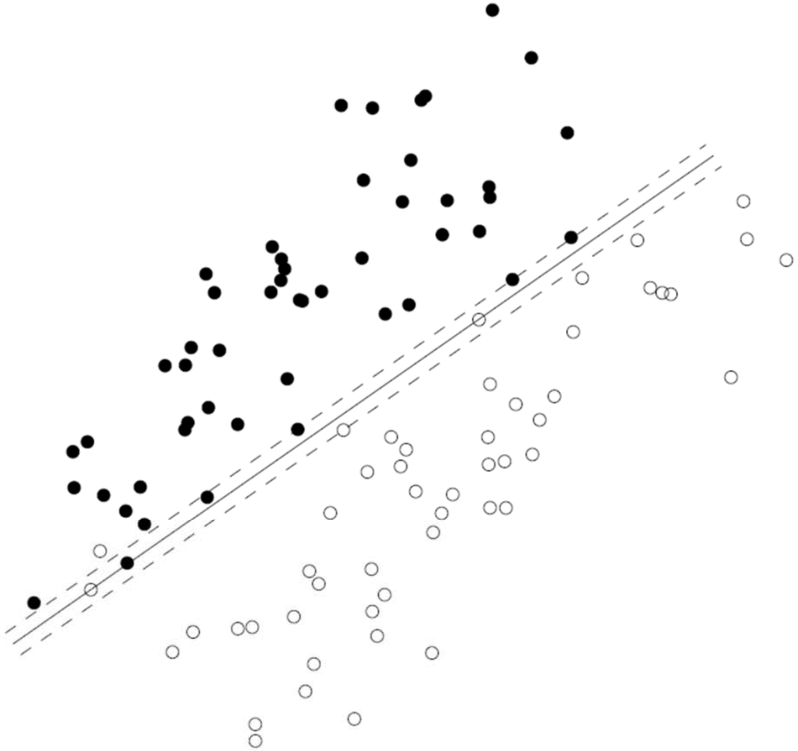


Figure 8.10 Approximate linear discrimination via linear programming. The points $x_1, \ldots, x_{50}$, shown as open circles, cannot be linearly separated from the points $y_1, \ldots, y_{50}$, shown as filled circles. The classifier shown as a solid line was obtained by solving the LP (8.25). This classifier misclassifies one point. The dashed lines are the hyperplanes $a^T z - b = \pm 1$. Four points are correctly classified, but lie in the slab defined by the dashed lines.

Support Vector Classifier

- More generally, we can consider the trade-off between the number of misclassified points, and the width of the slab $\{z -1 \leq a^{\top}z - b \leq 1\}$, which is given by $2/\|a\|_2$

$$
\begin{aligned}
\min \quad & \|a\|_2 + \gamma(1^{\top}u + 1^{\top}v) \\
\text{s.t.} \quad & a^{\top}x_i - b \geq 1 - u_i, i = 1, \ldots, N \\
& a^{\top}y_i - b \leq -(1 - v_i), i = 1, \ldots, M \\
& u \succcurlyeq 0,\ v \succcurlyeq 0
\end{aligned}
$$

  - We want to minimize the error and maximize the width of the slab and

Example

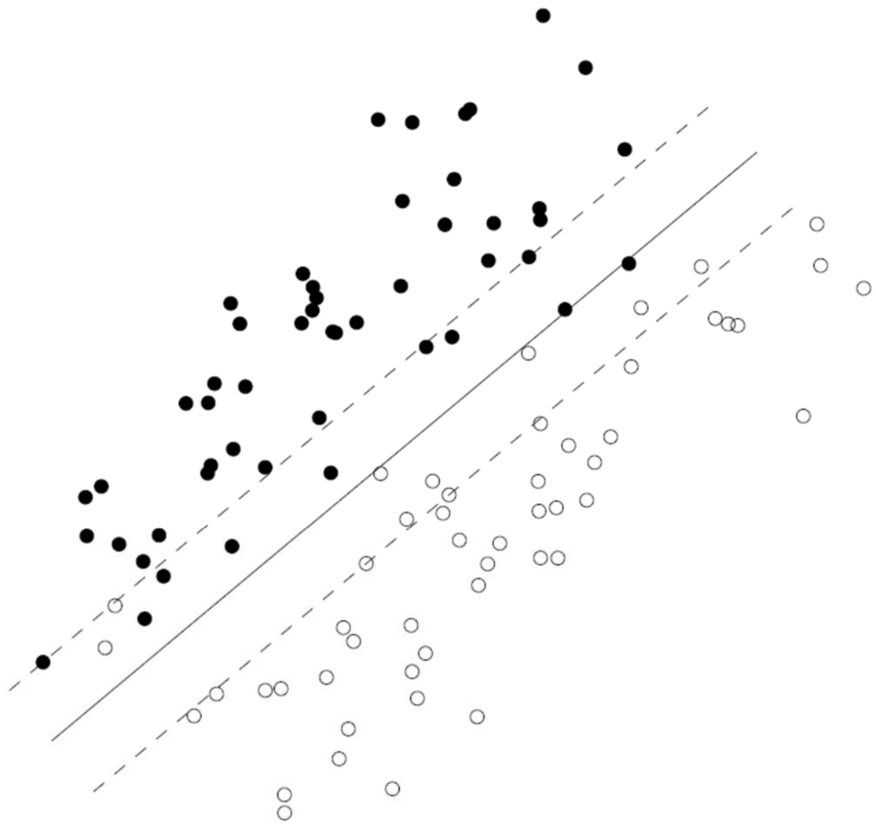


Figure 8.11 Approximate linear discrimination via support vector classifier, with $\gamma = 0.1$. The support vector classifier, shown as the solid line, misclassifies three points. Fifteen points are correctly classified but lie in the slab defined by $-1 < a^T z - b < 1$, bounded by the dashed lines.

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression

Logistic Regression

- z is a random variable with values 0 or 1, with a distribution that depends on $u \in \mathbf{R}^n$
  - Logistic Model

$$\text{prob}(z = 1) = \frac{\exp(a^\top u - b)}{1 + \exp(a^\top u - b)}$$

$$\text{prob}(z = 0) = \frac{1}{1 + \exp(a^\top u - b)}$$

- Given sets of points, $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_M\}$, arise as samples from the logistic model

Logistic Regression

Maximum Likelihood Estimation

$$\min \; -l(a,b)$$

- l is the log-likelihood function

$$l(a,b) = \sum_{i=1}^{N} (a^\top x_i - b) - \sum_{i=1}^{N} \log(1 + \exp(a^\top x_i - b)) - \sum_{i=1}^{M} \log(1 + \exp(a^\top y_i - b))$$

- If the two sets of points can be linearly separated, then the optimization problem is unbounded below
  - ✓ Add domain constraints

Example

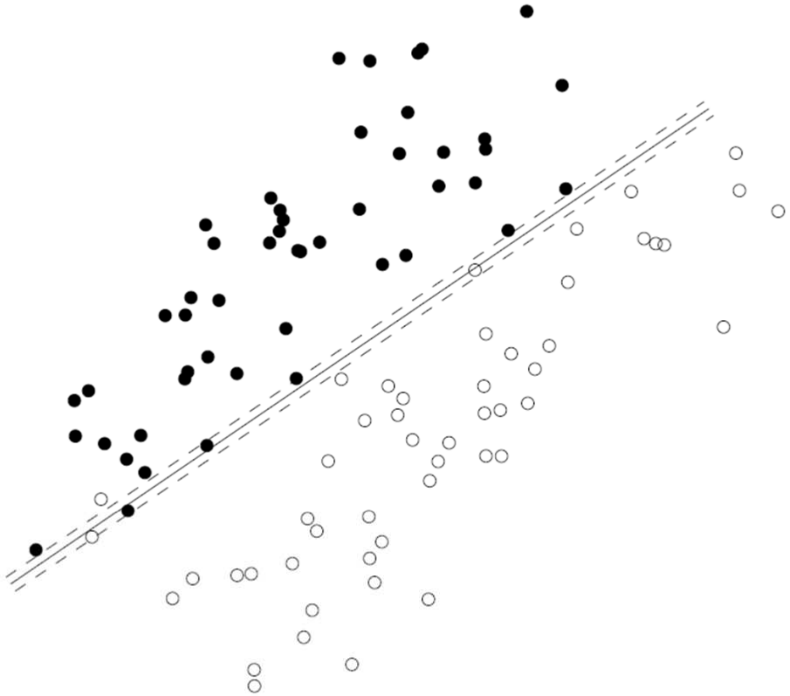


Figure 8.12 Approximate linear discrimination via logistic modeling. The points $x_1, \ldots, x_{50}$, shown as open circles, cannot be linearly separated from the points $y_1, \ldots, y_{50}$, shown as filled circles. The maximum likelihood logistic model yields the hyperplane shown as a dark line, which misclassifies only two points. The two dashed lines show $a^T u - b = \pm 1$, where the probability of each outcome, according to the logistic model, is 73%. Three points are correctly classified, but lie in between the dashed lines.

Summary

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Classification
  - Linear Discrimination
  - Support Vector Classifier
  - Logistic Regression