Quadratische Programmierung

Quadratische Optimierung unter
Nebenbedingungen und ihre
Anwendung mit SVM's

Inhalt

- Definitionen
- Allgemeiner (Nichtlinearer) Optimierer
- Quadratische Optimierung
- Umsetzung: Active Set Method
- Anwendbarkeit für SVM
- Sequential Minimal Optimization (SMO)
- Zusammenfassung
- Literatur

Einführung

Anwendungsgebiete der Optimierung

- Mathematik (Lösung von DGL, ...)
- Physik (Energieminimierung, Statik, ...)
- Wirtschaft (Portfolio-Optimierung, ...)
- Mustererkennung
- ...

Definitionen

Allgemeine (Nichtlineare) Optimierungsprobleme

Ein allgemeines Optimierungsproblem lässt sich folgendermaßen mathematisch formulieren:

$$\text{Minimiere} \qquad f(\vec{x}) \qquad \vec{x} \in E^n$$

$$\text{Unter Nebenbedingungen} \quad \left.\begin{array}{c} h_1(\vec{x})=0 \\ \vdots \\ h_m(\vec{x})=0 \end{array}\right\} \Leftrightarrow \vec{h}(\vec{x})=\vec{0} \quad \left.\begin{array}{c} g_1(\vec{x})\le 0 \\ \vdots \\ g_r(\vec{x})\le 0 \end{array}\right\} \Leftrightarrow \vec{g}(\vec{x})\le\vec{0}$$

$f(\vec{x})$ ist die zu minimierende **Zielfunktion**

$h_i(\vec{x})=0$ sind sog. **Gleichheitsbedingungen**

$g_j(\vec{x})\le 0$ sind sog. **Ungleichheitsbedingungen**

Definitionen

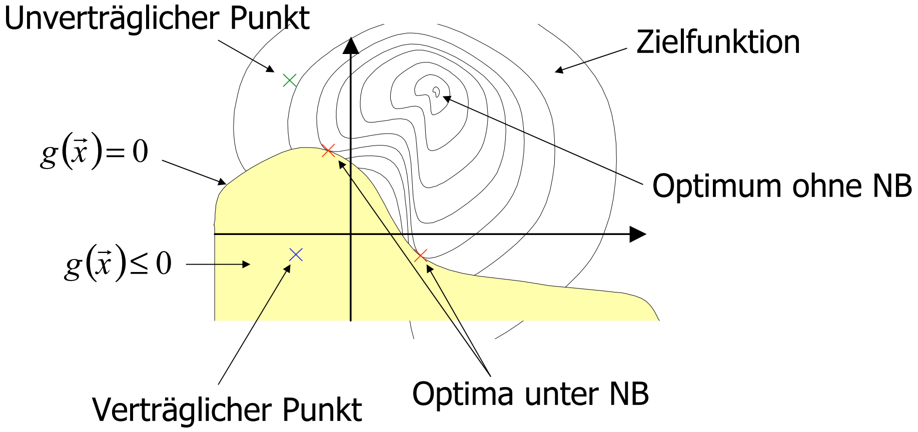
Verträglichkeit (*Feasibility*) von Vektoren & Aktivität von Nebenbedingungen

- Ein Vektor $\vec{x} \in E^n$ ist **verträglich**, wenn er alle Nebenbedingungen erfüllt.

- Eine Nebenbedingung $g_i(\vec{x}) \leq 0$ ist **aktiv** für einen verträglichen Vektor $\vec{x} \in E^n$ genau dann, wenn gilt $g_i(\vec{x}) = 0$.

Definitionen

Verträglichkeit & Aktive Nebenbedingungen



Definitionen

Tangentialebenen auf den Nebenbedingungen

- Für jeden verträglichen Vektor $\vec{x} \in E^n$ definieren seine aktiven Nebenbedingungen **Oberflächen** im Raum. Sind diese differenzierbar ist die zugehörige **Tangentialebene** durch $\vec{x}$ definiert als:

$$M = \left\{ \vec{y} \;\middle|\; \vec{\nabla}\vec{h}(\vec{x})\vec{y} = 0 \wedge \vec{\nabla}\vec{g}_{aktiv}(\vec{x})\vec{y} = 0 \right\}$$

Nichtlineare Optimierung

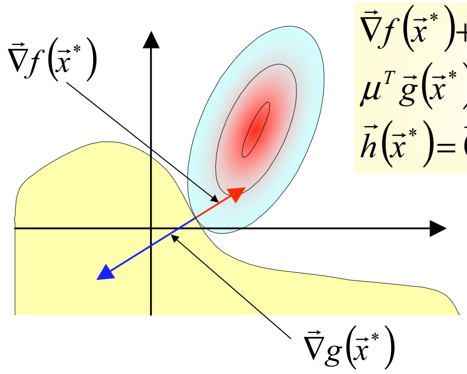
Allgemeines Vorgehen

Eine iterative Optimierung läuft im Allgemeinen folgendermaßen ab:

- Starte in einem Punkt der alle Nebenbedingungen erfüllt
- **Ermittle eine Richtung**, entlang derer eine sog. *Abstiegsfunktion* (z.B. Gradient) minimal und negativ ist. Gibt es keine solche Richtung: STOP
- **Bestimme entlang dieser Richtung das Minimum** (*line-search*) unter Berücksichtigung der Nebenbedingungen (1D-Optimierung)
- **Wiederhole** die letzten beiden Schritte

Lagrangeansatz
Lagrangemultiplikatoren

Allgemeine Lagrange-Bedingungen:



$$\vec{\nabla} f\left(\vec{x}^*\right) + \vec{\lambda}^T \vec{\nabla} \vec{h}\left(\vec{x}^*\right) + \vec{\mu}^T \vec{\nabla} \vec{g}\left(\vec{x}^*\right) = 0$$
$$\mu^T \vec{g}\left(\vec{x}^*\right) = \vec{0} \quad und \quad \vec{\lambda}, \vec{\mu} \geq \vec{0}$$
$$\vec{h}\left(\vec{x}^*\right) = \vec{0} \quad \vec{g}\left(\vec{x}^*\right) \leq \vec{0}$$

Quadratische Optimierung
Grundlagen

Das quadratische Optimierungsproblem hat folgende Form:

$$\begin{array}{ll} Minimiere & \frac{1}{2}\vec{x}^T Q \vec{x} + \vec{c}^T \vec{x} \\ Unter\ Nebenbedingungen & \vec{a}_i^T \vec{x} = b_i \quad i \in E \\ & a_i^T \vec{x} \leq b_i \quad i \in I \end{array}$$

Dabei ist Q *symmetrisch* und *positiv semidefinit*.

I und E sind Indexmengen über Gleichheits- und Ungleichheitsbedingungen

Quadratische Optimierung
Eigenschaften

- Die quadratische **Zielfunktion ist konvex**, das heißt jedes Minimum der eingeschränkten Funktion ist globales Minimum
- Die Nebenbedingungen definieren ein **konvexes Gebiet**
- Die notwendige Bedingung für das globale Optimum mit Lagrange Multiplikatoren ist **linear**!

Quadratische Optimierung
Bedingungen für ein Optimum

Die Bedingungen für das Optimum (ohne Ungleichheitsbedingungen) lauten dann:

$$Q\vec{x} + A^T\vec{\lambda} + \vec{c} = \vec{0}$$
$$A\vec{x} - \vec{b} = \vec{0}$$

Man erhält ein LGS in (n+m) Unbekannten mit (n+m) Gleichungen

Quadratische Optimierung
Lösungsformel

Mit etwas Rechnerei erhält man folgende explizite Lösungsformel:

$$\vec{x} = Q^{-1}A^T\left(AQ^{-1}A^T\right)^{-1}\left(AQ^{-1}\vec{c}+\vec{b}\right)-Q^{-1}\vec{c}$$

Aber was machen wir mit den Ungleichheitsbedingungen?

Active Set Method
Idee

Die Idee ist einfach:

Betrachte nur Aktive Untermengen (Working sets) der Nebenbedingungen, dadurch reduziert sich das Problem auf den explizit lösbaren Fall.

Active Set Method

Vorbemerkung

Sei W_k ein Working set bestehend aus Gleichheitsbedingungen und $\vec{x}_k$ ein verträglicher Vektor, dann lässt sich folgendes Problem Lösen:

$$\begin{array}{ll} \textit{Minimiere} & \frac{1}{2}\vec{d}_k^{\,T} Q \vec{d}_k + (\vec{c} + Q\vec{x}_k)\vec{d}_k \\ \textit{unter} & \vec{a}_i^{\,T}\vec{d}_k = 0 \quad i \in W_k \end{array}$$

Active Set Method

Vorbemerkung 2

Die vorangegangene Optimierung bestimmt die Richtung des nächsten Schritts. Wie weit man nun gehen kann definiert folgende Formel:

$$\alpha_k = \min_{\vec{a}_i^{\,T}\vec{d}_k > 0} \left\{ 1, \frac{b_i - \vec{a}_i^{\,T}\vec{x}_k}{\vec{a}_i^{\,T}\vec{d}_k} \right\}$$

Active Set Method

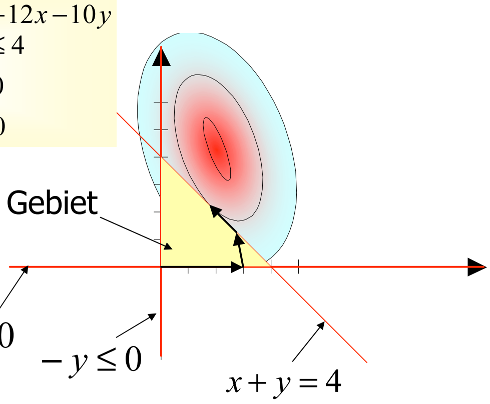
Algorithmus

- Bestimme die Richtung des nächsten Schritts
  - Falls $\vec{d}_k = \vec{0}$ gehe zu Schritt 3
- Setze $\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k$
  - Falls $\alpha_k < 1$ nimm die neue aktive Nebenbedingung in W_{k+1} auf.
  - Erhöhe k, Gehe zu Schritt 1
- Berechne den kleinsten Lagrangemultiplikator der Richtungsformel λ_q
  - Falls $\lambda_q \geq 0$ stop.
  - Ansonsten entferne die entspr. Nebenbedingung aus W_k und setze bei Schritt 1 fort

Active Set Method

Beispiel

$$\begin{array}{ll} \textit{Minimiere} & 2x^2 + xy + y^2 - 12x - 10y \\ & x + y \leq 4 \\ \quad\textit{unter} & -x \leq 0 \\ & -y \leq 0 \end{array}$$



Andere Methoden

Neben der Vorgestellten Active Set Method steht einem Entwickler eine ganze Palette von Alternativen zur Verfügung:

- (Modifizierte) *Gradientenverfahren*
- *Konjugierte Richtungen*
- (Quasi-) *Newton-Verfahren*
- ...

Anwendbarkeit für SVM

Eigenschaften der numerischen Verfahren

- Active Set Method
  - Speicherbedarf in $O(n^2)$ zur Speicherung der Matrix Q.
  - Hoher Rechenaufwand zur Bestimmung der Inversen im Richtungs-Such-Schritt
- Mögliche „Zusatzkosten“
  - Hessematrizen (Newton)
  - Eigenwerte und –vektoren (Konjugierte Richtungen

Anwendbarkeit für SVM

Wiederholung

SVM-spezifisches Quadratisches OP:

$$\text{Minimiere} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j K(\vec{x}_i, \vec{x}_j)\alpha_i \alpha_j - \sum_{i=1}^{N}\alpha_i$$

$$\text{unter} \quad 0 \le \alpha_i \le C \quad \forall i, \qquad \sum_{i=1}^{N} y_i \alpha_i = 0$$

Mit Entscheidungsfunktion:

$$u = \sum_{j=1}^{N} y_j \alpha_j K(\vec{x}_j, \vec{x}) + b$$

Anwendbarkeit für SVM

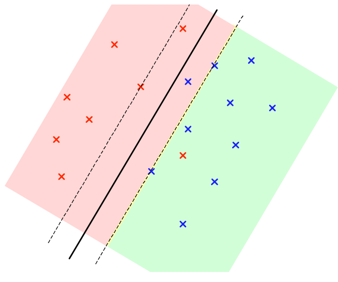
Hinreichende Bedingungen

Die Lösung ist optimal genau dann, wenn alle Trainingsbeispiele die sog. Karush-Kuhn-Tucker (KKT) Bedingungen erfüllen:

$$\alpha_i = 0 \Rightarrow y_i u_i \ge 1$$

$$0 < \alpha_i < C \Leftrightarrow y_i u_i = 1$$

$$\alpha_i = C \Rightarrow y_i u_i \le 1$$



Anwendbarkeit für SVM
Zahlenbeispiel

# Trainingsbeispiele	Speicherbedarf für Q
100	80 kB
1000	8 MB
2000	32 MB
5000	200 MB
10000	800 MB
20000	6.4 GB

Anwendbarkeit für SVM
Schlussfolgerung

Da numerische Optimierer die Matrix Q benötigen (möglicherweise noch wesentlich mehr), ist ihr Einsatz mit Supportvektormaschinen impraktikabel!

Spezielle Verfahren für SVM

Dekomposition

Dekomposition beschreibt die Zerlegung des Problems in Subprobleme kleinerer Dimension

Dies erfordert aber eine äußere Iteration, in der das Optimum des vollständigen Problems bestimmt wird

Spezielle Verfahren für SVM

SMO - Grundlagen

Osuna's Theorem

Ein großes quadratisches Problem kann in eine Folge von Subproblemen zerlegt werden, solange in jedem Schritt zumindest ein Beispiel, dass die KKT Bedingungen verletzt, hinzugenommen wird.

Spezielle Verfahren für SVM

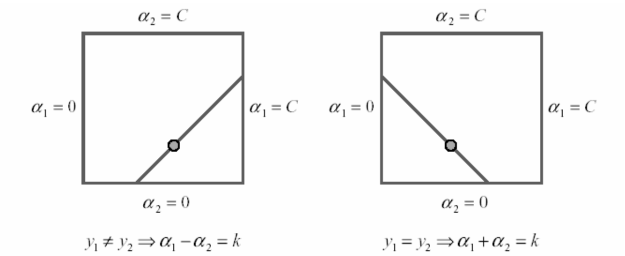
SMO – Sequential Minimal Optimization (Platt)

- Idee:
  - Löse eine Sequenz minimaler Subprobleme
    - Da die Gleichheitsbedingung immer erfüllt sein muss, müssen mindestens 2 Lagrangemultiplikatoren simultan optimiert werden
- Umsetzung:
  - Wähle heuristisch in jedem Schritt 2 Lagrangemultiplikatoren
  - Berechne analytisch die optimalen Werte dieser Multiplikatoren

Spezielle Verfahren für SVM

Subprobleme

Die zu lösenden Subprobleme haben folgende einfache Gestalt:



Zu ihrer Lösung benötigt man nur die beteiligten Samples!

Spezielle Verfahren für SVM
SMO - Eigenschaften

- Beliebig große Trainingsmengen können trainiert werden
- Keine komplizierten Optimierungen
  - Schneller
  - Genauer
- Sehr viele Iterationen ($O(n^2)$)

  (Das gleicht sich durch die schnelle Berechnung innerhalb einer Iteration wieder aus)

Zusammenfassung

- Optimierung ist ein iterativer Prozess, der in (kleinen) Schritten gegen das Optimum konvergiert
- Konventionelle Verfahren haben sehr hohen Speicheraufwand und sind für große Probleme impraktikabel
- SVM nutzen die Zerlegbarkeit des zugrundeliegenden quadratischen Problems in verschiedenen Implementationen verschieden gut aus

Literatur

Zur allgemeinen numerischen Optimierung:

David G. Luenberger ***Linear and Nonlinear Programming***
2nd Edition ***Addison-Wesley (1984)***

Zu den speziellen SVM Verfahren:

John C. Platt ***Technical Report MSR-TR-98-14***
Sequencial Minimal Optimization

Lagrange

Idee

Voraussetzung: f, g und h stetig differenzierbar

- Idee:
  - Betrachtet man die Tangentialebene im Optimum, erkennt man, dass der Gradient der Zielfunktion senkrecht auf dieser Tangentialebene steht (Projektion des Optimums der Zielfunktion auf die Nebenbedingungen)
  - Daraus folgt, dass der Gradient der Zielfunktion eine Linearkombination der Gradienten der aktiven Bedingungen ist!

Lagrange
Bedingungen

- Daraus ergibt sich folgende notwendige Bedingung für ein lokales Optimum:

$$\vec{\nabla} f(\vec{x}^*) + \vec{\lambda}^T \vec{\nabla} \vec{h}(\vec{x}^*) + \vec{\mu}^T \vec{\nabla} \vec{g}(\vec{x}^*) = 0$$
$$\mu^T \vec{g}(\vec{x}^*) = \vec{0} \quad und \quad \vec{\lambda}, \vec{\mu} \geq \vec{0}$$
$$\vec{h}(\vec{x}^*) = \vec{0} \quad \vec{g}(\vec{x}^*) \leq \vec{0}$$

$\vec{\lambda}, \vec{\mu}$ sind die Koeffizientenvektoren

= Lagrangemultiplikatoren

Spezielle Verfahren für SVM
Chunking (Vapnik)

- Idee
  - Sparse Darstellung der Lösung ausnutzen
    - Da nur Supportvektoren Einfluss auf die Lösung haben, reduziere das Problem auf die Supportvektoren
- Umsetzung
  - Löse in jedem Schritt ein Teilproblem, dass alle potentiellen Supportvektoren des letzten Schritts enthält.
  - Entferne neu gefundene Nicht-Supportvektoren
  - Füge dafür neue Vektoren hinzu, die die KKT-Bedingungen (erheblich) verletzen

Spezielle Verfahren für SVM

Chunking - Nachteile

- Ist die Zahl der Supportvektoren für das Problem groß, ergibt sich das gleiche Problem wie zuvor
- Man benötigt in jedem Schritt eine aufwendige numerische Optimierung