

E2 - Proteine



12. Tag: Optimierung und Parameterschätzung

Ursula Kummer, Sven Sahle
Femke Mensonides, Irina Surovtsova, Jürgen Zobeley

Optimierung



Generell:
Sucht Minimum/Maximum

Mögliche Fragestellungen:

- Ich möchte die Konzentration von ATP im steady state möglichst klein oder gross haben
- Ich möchte einen möglichst großen Flux durch eine bestimmte Reaktion
- etc.

Optimierungsmethoden



Lokale Optimierer:

- Newton
- Steepest Descent
- Levenberg-Marquardt

Globale Optimierer:

- Stochastische Methoden
- Genetic Algorithms

Newton



Problem: Finde $f(x) = 0$

Ansatz: Entwickle Taylorreihe um Startpunkt x_0

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{1}{2}h^2f''(x_0) \dots\dots$$

Da $f(x_0 + h)$ gleich Null sein sollte

$$\Rightarrow h = -f(x_0)/f'(x_0)$$

Rechenvorschrift:

$$x = x_0 + h = x_0 - f(x_0)/f'(x_0)$$

$$\text{Iterativ: } x^{(k+1)} = x^{(k)} - f(x^{(k)})/f'(x^{(k)})$$

Newton



Aber: Eigentlich soll ja $f'(x) = 0$ gefunden werden.

Daher:

$$f'(x_0 + h) = f'(x_0) + hf''(x_0) + \frac{1}{2}h^2f'''(x_0) \dots\dots$$

$$\Rightarrow h = -f'(x_0)/f''(x_0)$$

Rechenvorschrift:

$$x = x_0 + h = x_0 - f'(x_0)/f''(x_0)$$

$$\text{Iterativ: } x^{(k+1)} = x^{(k)} - f'(x^{(k)})/f''(x^{(k)})$$

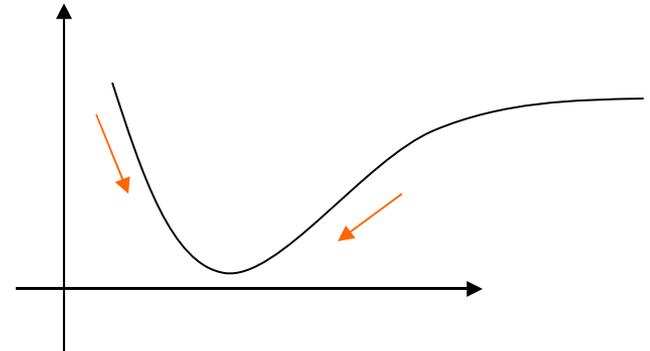
Newton



- Konvergiert quadratisch wenn nahe der Lösung
- Konvergiert garnicht wenn weit weg
- Neigt manchmal dazu, überzuschießen und überzukompensieren, was in Oszillationen mündet

Steepest Descent

- Verfolgt einfach den Gradienten (1. Ableitung) der Funktion
- Konvergiert linear und immer
- Sehr langsam
- Oft nur lokales Minimum
- Keine Möglichkeit, Variablen einzuschränken



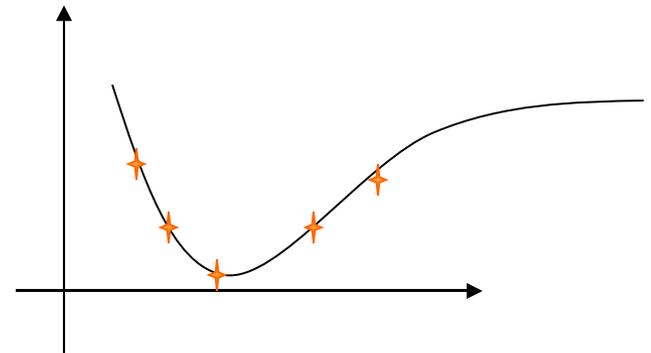
Levenberg-Marquardt



- Hybrid zwischen Newton und Steepest Descent
- When weit weg vom Minimum -> Steepest Descent
- Weit gebraucht und ziemlich mächtig

Zufallsuche - Random Search

- Wie der Name impliziert -> Zufällige Parameterwerte
-> Überprüfung der Funktion von Interesse
- Keine Konvergenz!
- Viel Geduld erforderlich
- Wird irgendwann das globale Optimum finden



Direkte Suche



- Kommt ohne Ableitungen aus
- Lokaler Optimierungsalgorithmus
- Benutzt heuristische Methode, um die Funktion zu erniedrigen, indem die letzten Iterationsschritte evaluiert werden

Hooke-Jeeves



- Direkte Suche
- Bei jedem Schritt wird ein „Muster“ von Punkten definiert,
indem jeder Parameter variiert wird und geschaut wird,
was die Optimierung verbessert.
- Das ganze „Muster“ wird dann transferiert
- Die neue Lage des Musters wird durch Extrapolation bestimmt
- Meist ziemlich effektiv

Genetische Algorithmen



„Strukturierte Zufallssuche“

- Selektion eines „Chromosoms“ = Parameterset
- Evaluation der Fitness
- Fittesten können sich paaren und „Kinder“ erzeugen

Hybrid GO-LO



- Um die Geschwindigkeit von lokalen Methoden (LO) mit der Möglichkeit der globalen Methoden (GO) zu verbinden
- Benutzt GO-Methoden (z.B. evolutionäre Algorithmen), um die Nähe des Minimums zu finden
- Benutzt LO-Methoden (z.B. Levenberg, Steepest Descent) um dieses genau zu bestimmen

Startwerte



- Gute Startwerte sind wichtig, um überhaupt und schnell das Optimum finden zu können
- Gute Startwerte sind z.B. Daten generiert aufgrund von in vitro-Daten, Sequenzvergleich etc.
- S. Praktikum 1. Woche

Parameterschätzung



- Gegeben sind experimentelle Werte
- Gesucht die Parameterwerte, die das Modell am besten an diese angleichen
- Keine Aussage, wie verlässlich die Schätzung ist
- Lineare Regression ist Spezialfall

Summe der Quadrate



Der Restbetrag (**residue**) ist die Differenz zwischen beobachtetem und vorhergesagtem Wert einer Funktion

Residue = beobachteter Wert - vorhergesagter Wert

Der Fehler aus der Summe der Quadrate

sum-of-squares error (SSE) wird berechnet wie folgt

$$\begin{aligned} \text{SSE} &= \text{Sum of squares of residues} \\ &= \text{Sum of } (y_{\text{observed}} - y_{\text{predicted}})^2 \end{aligned}$$

Um so kleiner SSE, um so besser ist der „Fit“ des Modells an die experimentellen Daten.