Advanced Machine Learning Subsidary Notes

Lecture 11: Constrained Optimisation

Adam Prügel-Bennett

April 21, 2020

1 Keywords

- Lagrangians, Inequalities, KKT, Linear Programming,

2 Main Points

2.1 Equality Constraints

- If we want to minimise $f(\boldsymbol{x})$ subject to the constraint $g(\boldsymbol{x}) = 0$ this is equivalent to solving the problem
$$\min_{\boldsymbol{x}} \max_{\alpha} \mathcal{L}(\boldsymbol{x}, \alpha)$$
where $\mathcal{L}(\boldsymbol{x}, \alpha)$ is a *Lagrangian* given by
$$\mathcal{L}(\boldsymbol{x}, \alpha) = f(\boldsymbol{x}) - \alpha\, g(\boldsymbol{x})$$
- α is a Lagrange multiplier that is determined by the joint optimisation problem
- Note that we seek a saddle-point
  - We minimise with respect to $\boldsymbol{x}$ and maximise with respect to α
  - We can't escape this
    - if we multiply α by -1 we just changing the directions of the α axis but still have a saddle-point
    - if we multiply both terms by -1 then we would end up minimising with respect to α, but maximising with respect to $\boldsymbol{x}$
- The solution to our problem must satisfy
$$\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{x}, \alpha) = \boldsymbol{\nabla} f(\boldsymbol{x}) - \alpha\, \boldsymbol{\nabla} g(\boldsymbol{x}) = 0, \qquad\qquad \frac{\partial \mathcal{L}(\boldsymbol{x}, \alpha)}{\partial \alpha} = g(\boldsymbol{x}) = 0$$
  - The second equation ensures that we sit on the constraint
  - The first equation says that the gradient of $f(\boldsymbol{x})$ must be perpendicular to the constraint
  - This is necessary for the solution to be a (local) minimum (i.e. we can not get an improvement by moving along the constraint)
  - There can be multiple solutions: these equations at a satisfied for any local minima on the constraint
- If we have multiple equality constraints we just use multiple Lagrange multipliers

2.2 Inequality constraints

- If we are minimising $f(\boldsymbol{x})$ subject to an inequality constraint $g(\boldsymbol{x}) \geq 0$ then one of two things can happen

  1. Either we have a (local) minimum of $f(\boldsymbol{x})$ that satisfies the constraint or
  2. We have a local minimum on the constraint

- We can therefore solve this problem by again using a Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \alpha) = f(\boldsymbol{x}) - \alpha\, g(\boldsymbol{x})$$

  with the additional constraint $\alpha \geq 0$

  - If the minimum lies in a region that satisfies the constraint then we just set $\alpha = 0$ and minimise $f(\boldsymbol{x})$
  - If the solution lies on the constraint we again have $\boldsymbol{\nabla} f(\boldsymbol{x}) = \alpha\, \boldsymbol{\nabla} g(\boldsymbol{x})$, but now $\alpha > 0$ which means that not only is there no improving direction along the constraint, but also the negative-gradient of $f(\boldsymbol{x})$ points in the direction where $g(\boldsymbol{x})$ becomes smaller, i.e. in the region that violates the constraint
    * note if $\alpha < 0$ we could find a better solution moving away from the constraint into the feasible region
  - *Karush-Kuhn-Tucker (KKT) Conditions*
    * For inequality constraints we require either
      1. $\alpha = 0$ and there is an unconstrained minimum in the regions $g(\boldsymbol{x}) \geq 0$ or
      2. $\alpha > 0$ and the solution lies on $g(\boldsymbol{x}) = 0$

- If we have multiple inequality constraints we just introduce a Lagrange multiply for each constraint with $\alpha \geq 0$

2.3 Duality

- Our problem of solving $f(\boldsymbol{x})$ subject to some constraints is known as the *primal problem*

- If our problem is sufficiently simple we can sometimes find a solution $\boldsymbol{x}^*(\boldsymbol{\alpha})$ that satisfies

$$\boldsymbol{\nabla} \mathcal{L}(\boldsymbol{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) = 0$$

- This leaves us with the *dual problem*

$$\max_{\boldsymbol{\alpha}} \; \mathcal{L}(\boldsymbol{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha})$$

  possible with constraints on $\boldsymbol{\alpha}$ (e.g. $\alpha_i \geq 0$) that arise from KKT conditions

- **Linear Programming**

  - In linear programming we are trying to find a value of $\boldsymbol{x}$ that minimises a linear objective function $\boldsymbol{c}^{\mathsf{T}} \boldsymbol{x}$ subject to linear constraints $\mathbf{M}\, \boldsymbol{x} = \boldsymbol{b}$ (and/or $\mathbf{M}\boldsymbol{x} \geq \boldsymbol{b}$)
  - We can write this as a Lagrange problem

$$\mathcal{L} = \boldsymbol{c}^{\mathsf{T}} \boldsymbol{x} - \boldsymbol{\alpha}^{\mathsf{T}} \left(\mathbf{M}\boldsymbol{x} - \boldsymbol{b} \right)$$

    (subject to constraints $\boldsymbol{\alpha} \geq 0$ if we have inequality constraints in the primal problem)
  - We can rearrange the Lagrangian as

$$\mathcal{L} = \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{b} + \left(\boldsymbol{c}^{\mathsf{T}} - \boldsymbol{\alpha}^{\mathsf{T}} \mathbf{M} \right) \boldsymbol{x}$$

- Using the identity $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{b} = \boldsymbol{b}^{\mathsf{T}}\boldsymbol{a}$ w can rewrite this as

$$\mathcal{L} = \boldsymbol{b}^{\mathsf{T}}\boldsymbol{\alpha} - \boldsymbol{x}^{\mathsf{T}}\left(\mathbf{M}^{\mathsf{T}}\boldsymbol{\alpha} - \boldsymbol{c}\right)$$

- But we can now treat $\boldsymbol{x}$ as a Lagrange multiplier so we get the *dual problem*

$$\max_{\boldsymbol{\alpha}}\ \boldsymbol{b}^{\mathsf{T}}\boldsymbol{\alpha}$$

subject to the constraint

$$\mathbf{M}^{\mathsf{T}}\boldsymbol{\alpha} = \boldsymbol{c}$$

- If the original constraints were inequality constraints the $\boldsymbol{\alpha} \geq 0$
- The dimensionality of the dual problem can sometimes be much lower than that of the primal problem making it easier to solve

- **Quadratic Program**
  - In a quadratic program we have to minimise a quadratic loss $\boldsymbol{x}^{\mathsf{T}}\mathbf{Q}\,\boldsymbol{x}$ subject to linear constraints $\mathbf{M}\,\boldsymbol{x} = \mathbf{b}$ (or $\mathbf{M}\,\boldsymbol{x} \geq \mathbf{b}$)
  - For there to be a unique minimum $\mathbf{Q}$ must be positive definite (which is sometimes written $\mathbf{Q} \succ 0$)
  - We can write a Lagrangian

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\alpha}) = \boldsymbol{x}^{\mathsf{T}}\mathbf{Q}\,\boldsymbol{x} - \boldsymbol{\alpha}^{\mathsf{T}}\left(\mathbf{M}\,\boldsymbol{x} - \boldsymbol{b}\right)$$

  - The solution is given by $\max\limits_{\boldsymbol{\alpha}}\min\limits_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\alpha})$
  - If the constraints are inequality constraints then $\alpha_i \geq 0$
  - The minimum with respect to $\boldsymbol{x}$ is given by

$$\boldsymbol{\nabla}_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\alpha}) = 2\,\mathbf{Q}\,\boldsymbol{x} + \mathbf{M}^{\mathsf{T}}\boldsymbol{\alpha} = 0$$

  - So that $\boldsymbol{x}^* = \frac{1}{2}\mathbf{Q}^{-1}\mathbf{M}^{\mathsf{T}}$
  - Substituting this back into the Lagrangian we get the dual problem

$$\max_{\boldsymbol{\alpha}} -\frac{1}{4}\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{M}\mathbf{Q}^{-1}\mathbf{M}^{\mathsf{T}}\boldsymbol{\alpha} + \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{b}$$

  with $\alpha_i \geq 0$ if we started with inequality constraints
    * note in the derivation that we end up with two terms proportional to $\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{M}\mathbf{Q}^{-1}\mathbf{M}^{\mathsf{T}}\boldsymbol{\alpha}$ one partially cancelling the other

3 Exercises

3.1 Quadratic with a linear constraint

- Consider minimising $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathsf{T}}\mathbf{Q}\boldsymbol{x}$ subject to the constraint $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x} = b$
  1. Write a Lagrangian for this problem
  2. Find the minimum of the Lagrangian with respect to $\boldsymbol{x}$
  3. Write down and solve the dual problem
  4. Hence write down a solution to the primal problem
- See answers, but also experiments

3.2 Saddle Point

- Strangely (for me at least) the optimum of a constrained optimisation problem is given by the saddle-point of the Lagrangian
- Consider the problems of minimising $x^2/2$ subject to the constraint $x = 1$
  1. Write down the Lagrangian
  2. Calculate the Hessian matrix (matrix of second derivatives)
  3. Compute the eigenvalues of the Hessian (show that they have different signs everywhere so there are no maxima or minima)
- See answers

4 Experiments

4.1 Quadratic with a linear constraint

- Let X be a 10×5 random matrix with elements drawn from $\mathcal{N}(0, 1)$
- Let $\mathbf{Q} = \mathbf{X}^\mathsf{T}\mathbf{X}$
  - Check that this is positive definite
- Let $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\mathsf{T}\mathbf{Q}\boldsymbol{x}$
- Let $\boldsymbol{a}$ be a random vector with 5 elements drawn from $\mathcal{N}(0, 1)$
- We want to minimise $f(\boldsymbol{x})$ subject to the constraint $\boldsymbol{a}^\mathsf{T}\boldsymbol{x} = 1$
- Work out the Lagrangian, $L(\boldsymbol{x}, \alpha)$ for this system
- Write an iterative gradient solver that
  1. Makes steps $\boldsymbol{x} \leftarrow \boldsymbol{x} - r\boldsymbol{\nabla} L(\boldsymbol{x}, \alpha)$
  2. Makes steps $\alpha \leftarrow \alpha + r\frac{\partial L(\boldsymbol{x},\alpha)}{\partial}\alpha$
- Note you will have to tune the learning step r
- Compare the solution you find by running your algorithm until convergence with the exact result (see exercise and/or answer)

```
function [x,alpha] = optimise(Q,a,r)
  x = rand(5,1);                                % initialise x
  alpha = 0;                                      % initialise alpha
  for t=1:1000
    x = x - r*(Q*x-alpha*a);                  % x = x - r  \grad L
    alpha = alpha + r*(-(a'*x-1));          % alpha = alpha + r dL/d\alpha
  endfor
endfunction

[x,alpha] = optimise(Q,a,0.1)

alphaTheory = 1/(a'*Q*a)
xTheory = inv(Q)*a*alphaTheory
```

5 Answers

5.1 Quadratic with a linear constraint

1. The Lagrangian is given by

$$\mathsf{L}(\boldsymbol{x}, \alpha) = \frac{1}{2}\,\boldsymbol{x}^{\mathsf{T}}\mathbf{Q}\boldsymbol{x} - \alpha\,(\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x} - b)$$

2. Minimising with respect to $\boldsymbol{x}$ we get

$$\boldsymbol{\nabla}\mathsf{L}(\boldsymbol{x}, \alpha) = \mathbf{Q}\,\boldsymbol{x} + \alpha\,\boldsymbol{a} = 0$$

   or $\boldsymbol{x} = \alpha\,\mathbf{Q}^{-1}\boldsymbol{a}$

3. Thus the dual problem is

$$\max_{\alpha} -\frac{1}{2}\alpha^2\,\boldsymbol{a}^{\mathsf{T}}\,\mathbf{Q}^{-1}\boldsymbol{a} + \alpha\,b$$

   - The solution to the dual problem is

$$\alpha = \frac{b}{\boldsymbol{a}^{\mathsf{T}}\,\mathbf{Q}^{-1}\boldsymbol{a}}$$

4. Thus the solution to the primal problem is

$$\boldsymbol{x} = \frac{b\,\mathbf{Q}^{-1}\boldsymbol{a}}{\boldsymbol{a}^{\mathsf{T}}\,\mathbf{Q}^{-1}\boldsymbol{a}}$$

   - Note that in most quadratic programming problems we are dealing with many inequality constraints so solving the dual problem in closed form isn't necessarily easy

5.2 Saddle Point

- Just do it
  1. The Lagrangian is given bye

$$\mathcal{L} = \frac{x^2}{2} - \alpha\,(x-1)$$

  2. The Hessian is given by

$$\mathbf{H} = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$$

  3. The traces $T = 1$ and the determinant $D = -1$ So that

$$\lambda = \frac{T \pm \sqrt{T^2 - 4\,D}}{2} = \frac{1 \pm 5}{2} = \{1.618, -0.618\}$$

     If you prefer you can compute the eigenvalues numerically

- Note that whatever we do the determinant will be negative leading to a negative eigenvalue (the determinant is equal to the product of eigenvalues). This would be true if we were maximising $-x^2/2$. You can change the constraints or the objective function, but you will still get eigenvalues of different signs.