CSE250A Fall ’12: Discussion Week 2

Aditya Menon (akmenon@ucsd.edu)

October 8, 2012

1 Administration

TAs: Phuc Nguyen, Aditya Menon (akmenon@ucsd.edu), Bryan Lunt

Office hours:
Phuc: Tuesday 3 - 4pm, EBU3B B240A
Aditya: Wednesday 11am - 12pm, EBU3B B240A
Bryan: Friday 12pm - 1pm, EBU3B B240A

2 Announcements

Homework 1 is due on Thursday.

3 Schedule for today

- Conditional probability.
- Bayes rule.
- Lagrange multipliers.
- Questions?

4 Conditional probability

If A, B are random variables, then the conditional probability of A given B is

$$\Pr[A|B] = \frac{\Pr[A, B]}{\Pr[B]}.$$

By implied universality, this is equivalent to saying

$$\Pr[A = a|B = b] = \frac{\Pr[A = a, B = b]}{\Pr[B = b]}$$

for all appropriate values of a, b. If A, B are independent, then it is true that

$$\Pr[A, B] = \Pr[A] \Pr[B],$$

so that

$$\Pr[A|B] = \frac{\Pr[A,B]}{\Pr[B]} = \Pr[A].$$

Think about conditional probability as a Venn diagram. Suppose we perform n random trials of an experiment, and measure the occurrence of events A and B. Let U be the universe of all possible outcomes, so that $|U| = n$. Note that the joint probability is

$$\Pr[A,B] = \frac{|A \wedge B|}{|U|} = \frac{|A \wedge B|}{|A|} \cdot \frac{|A|}{|U|} = \Pr[B|A]\Pr[A].$$

Example. Applicants to UCSD are either instant-accepts (1%), instant-rejects (9%), or require analysis by GradCom (90%). An initial pass of applications can remove all instant rejects. What is the probability that an application remaining after this pass requires analysis?

$$\Pr[S = \text{Analysis}|S \neq \text{Reject}] = \frac{\Pr[S = \text{Analysis}, S \neq \text{Reject}]}{\Pr[S \neq \text{Reject}]} = \frac{0.90}{0.91} \approx 0.9890.$$

5 Bayes' rule

Bayes' rule states that

$$\Pr[A|B] = \frac{\Pr[B|A]\Pr[A]}{\Pr[B]} = \frac{\Pr[A,B]}{\Pr[B]}.$$

From a Venn diagram perspective, we know that

$$\Pr[A|B] = \frac{|A \wedge B|}{|B|} = \frac{|A \wedge B|/|U|}{|B|/|U|} = \frac{\Pr[A,B]}{\Pr[B]}$$

Example. Of all applicants to UCSD, 1% are academically brilliant. If an applicant is academically brilliant, UCSD admits the student 95% of the time. If not, UCSD admits the student 20% of the time. Elton gets admitted to UCSD. What is the probability that he is academically brilliant?

Let A denote admittance, B brilliance. Then,

$$\Pr[B = 1] = 0.01$$

$$\Pr[A = 1|B = 1] = 0.95$$

$$\Pr[A = 1|B = 0] = 0.20$$

We want to know

$$\Pr[B = 1|A = 1] = \frac{\Pr[A = 1|B = 1]\Pr[B = 1]}{\Pr[A = 1]} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.20 \cdot 0.99} \approx 0.0458.$$

6 Lagrange multipliers

To minimize a constraint objective such as

$$\min_x f(x) : g(x) = b,$$

form the Lagrangian function

$$\mathcal{L}(x, \lambda) = f(x) + \lambda(g(x) - b).$$

Setting $\partial\mathcal{L}/\partial\lambda = 0$ we get that $g(x^*) = b$. This means that at a stationary point of $\mathcal{L}$, the constraint holds. To find the value of x^*, we set $\partial\mathcal{L}/\partial x = 0$, and solve for (x^*, λ^*) using this equation and the fact that $g(x^*) = b$.

Aside. Formally, it is the case that $\sup_{\lambda \geq 0} \inf_x \mathcal{L}(x, \lambda) = \inf_{x: g(x)=b} f(x)$ when strong duality holds.

Example. Consider

$$\min_{x,y} 2xy : x^2 + y^2 = 1.$$

Form the Lagrangian:

$$\mathcal{L}(x, y, \lambda) = 2xy + \lambda(1 - (x^2 + y^2)).$$

Compute the derivative wrt x:

$$\frac{\partial\mathcal{L}}{\partial x} = 2y - 2\lambda x = 2 \cdot (y - \lambda x),$$

and then wrt y:

$$\frac{\partial\mathcal{L}}{\partial y} = 2x - 2\lambda y = 2 \cdot (x - \lambda y).$$

At optimality, the partial derivatives tell us

$$\begin{aligned} \lambda^* x^* &= y^* \\ \lambda^* y^* &= x^* \\ (x^*)^2 + (y^*)^2 &= 1. \end{aligned}$$

So, $\lambda^* = \pm 1$. It then follows that $(x^*, y^*) = 1/\sqrt{2} \cdot (\pm 1, \pm 1)$. But let's go back to the objective. Observe that $x^* y^*$ is minimized when they possess different signs. So, $\lambda^* = -1$, and $(x^*, y^*) = (1/\sqrt{2}, -1/\sqrt{2})$ or $(-1/\sqrt{2}, 1/\sqrt{2})$.

Generalizing the example. The matrix generalization is the problem

$$\min_x x^T A x : ||x||_2 = 1$$

where A is symmetric. The Lagrangian is

$$\mathcal{L}(x, \lambda) = x^T A x + \lambda(1 - ||x||_2).$$

It is easy to check that

$$\frac{\partial\mathcal{L}}{\partial x} = (A + A^T)x - 2\lambda x = 2(Ax - \lambda x).$$

At optimality,

$$Ax^* = \lambda^* x^*$$

so that x^* is an eigenvector of A, and λ^* the eigenvalue. The value of the objective is

$$f(x^*) = \lambda^*$$

and so the solution is to pick the eigenvector corresponding to the smallest eigenvalue.