Recitation #5

10-606

September 2025

Convexity and Optimization

1. Let $f(x) = x^2 + 3x + 5$. Is f convex? **Solution:** Yes. The second derivative of f is $f''(x) = 2$, which is strictly positive so f is convex.

2. Compute the Hessian of the function $f(x, y) = x^2 + y^2 + xy$. Can we conclude whether or not f is convex? **Solution:** The Hessian matrix of $f(x, y)$ is:

$$H(f) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

   The eigenvalues of $H(f)$ are 3 and 1, both positive, so $f(x, y)$ is convex.

3. Show that the Hessian of $f(x) = x^\top Ax$ is $A + A^\top$. **Solution:** The Hessian is the Jacobian of the gradient. The gradient is $(A + A^\top)x$. The partial derivative of the i-th entry with respect to x_j is $a_{ij} + a_{ji}$, which is precisely the (i, j)-th entry of $A + A^\top$.

4. Consider the loss function in linear regression $L(\beta) = \|y - X\beta\|^2$. Prove that this loss function is convex. **Solution:** The loss function $L(\beta) = (y - X\beta)^T(y - X\beta)$ is quadratic in β, and its Hessian is $H(L) = 2X^TX$. Since X^TX is positive semi-definite (why?), the Hessian is positive semi-definite, proving that the loss function is convex.

5. Perform two iterations of gradient descent for $f(x) = x^2 + 4x + 4$, starting at $x_0 = 3$, with a learning rate $\alpha = 0.1$. **Solution:** The gradient is $\nabla f(x) = 2x + 4$. At $x_0 = 3$, we have:

$$\nabla f(3) = 2(3) + 4 = 10.$$

   The first iteration gives:

$$x_1 = x_0 - \alpha \nabla f(x_0) = 3 - 0.1 \times 10 = 2.$$

   The gradient at $x_1 = 2$ is:

$$\nabla f(2) = 2(2) + 4 = 8.$$

The second iteration gives:

$$x_2 = x_1 - \alpha \nabla f(x_1) = 2 - 0.1 \times 8 = 1.2.$$

After two iterations, $x_2 = 1.2$.

6. Consider the function $f(x, y) = x^2 + y^2$. Starting at $(x_0, y_0) = (1, 2)$, perform one iteration of gradient descent with learning rate $\alpha = 0.1$. **Solution:**
The gradient of $f(x, y) = x^2 + y^2$ is $\nabla f = (2x, 2y)$. At $(1, 2)$, the gradient is $(2, 4)$. Updating using gradient descent:

$$x_1 = x_0 - \alpha \frac{\partial f}{\partial x} = 1 - 0.1 \times 2 = 0.8,$$

$$y_1 = y_0 - \alpha \frac{\partial f}{\partial y} = 2 - 0.1 \times 4 = 1.6.$$

Thus, the new point is $(x_1, y_1) = (0.8, 1.6)$.

Probability

1. You roll a fair six-sided die.

   (a) What is the sample space?

   (b) Define the event "rolling an even number." What is its probability?

   (c) Define the random variable $X =$ (outcome mod 2). Give the distribution of X.

   Solution. (a) $\Omega = \{1, 2, 3, 4, 5, 6\}$. (b) $E = \{2, 4, 6\}$, so $P(E) = 3/6 = 1/2$. (c) X takes values $\{0, 1\}$. $P(X = 0) = P(\text{even}) = 1/2$, $P(X = 1) = 1/2$.

2. A fair coin is flipped 3 times.

   (a) What is the probability of exactly two heads?

   (b) What is the probability of at least one head?

   (c) If the coin were biased with $P(H) = 0.7$, what changes?

   Solution. (a) $\binom{3}{2}(1/2)^3 = 3/8$. (b) $1 - P(0 \text{ heads}) = 1 - (1/2)^3 = 7/8$. (c) For $p = 0.7$, let $E_{i,j}$ be the event "toss i and toss j are the only heads." Then $P(E_{i,j}) = p^2(1-p)$. So $P(\text{exactly two heads}) = P(E_{1,2}) + P(E_{2,3}) + P(E_{1,3}) = 3p^2(1-p)$. Likewise $P(\text{ at least one head}) = 1 - P(\text{no heads}) = 1 - (1-p)^3 = 1 - 0.3^3 = 0.973$.

3. Two fair dice are rolled. Let X=value of die 1, Y=value of die 2.

   (a) What is $P(X = 3, Y = 5)$?

(b) What is $P(X = 3 \mid X + Y = 8)$?

(c) Are X and Y independent?

Solution. (a) Each outcome equally likely with prob 1/36, so 1/36. (b) Outcomes with sum 8: (2,6),(3,5),(4,4),(5,3),(6,2). 5 equally likely. Only one has $X = 3$. So 1/5. (c) Independence: for any a, b, $P(X = a, Y = b) = 1/36 = P(X = a)P(Y = b)$. Yes, they are independent.