CPSC-540 Probabilistic Machine Learning 2011

Homework # 1
Due Thursday, Th 27 12:30pm.

NAME:____

Signature:____

STD. NUM: _______________________________________

General guidelines for homeworks:

You are encouraged to discuss the problems with others in the class, but all write-ups are to be done on your own.
Homework grades will be based not only on getting the "correct answer," but also on good writing style and clear presentation of your solution. It is your responsibility to make sure that the graders can easily follow your line of reasoning.
Try every problem. Even if you can't solve the problem, you will receive partial credit for explaining why you got stuck on a promising line of attack. More importantly, you will get valuable feedback that will help you learn the material.
Please acknowledge the people with whom you discussed the problems and what sources you used to help you solve the problem (e.g. books from the library). This won't affect your grade but is important as academic honesty.
When dealing with python exercises, please attach a printout with all your code and show your results clearly.

1 Gaussian decision boundaries

Let $p(x|y=j) = \mathcal{N}(x|\mu_j, \sigma_j)$ where $j = 1, 2$ and $\mu_1 = 0, \sigma_1^2 = 1, \mu_2 = 1, \sigma_2^2 = 10^6$. Let the class priors be equal, $p(y=1) = p(y=2) = 0.5$.

1. Find the decision region

   $$R_1 = \{x : p(x|\mu_1, \sigma_1) \geq p(x|\mu_2, \sigma_2)\} \tag{1}$$

   Sketch the result. Hint: draw the curves and find where they intersect. Find *both* solutions of the equation

   $$p(x|\mu_1, \sigma_1) = p(x|\mu_2, \sigma_2) \tag{2}$$

   Hint: recall that to solve a quadratic equation $ax^2 + bx + c = 0$, we use

   $$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3}$$

2. Now suppose $\sigma_2 = 1$ (and all other parameters remain the same). What is R_1 in this case?

3. In both cases, use Matlab's symbolic algebra toolbox to confirm your answers (Hint: you'll need the commands `normpdf` and `solve`). Hand in the scripts and the plots of the densities illustrating where they intersect.

2 Dummy encoding and linear models

Consider a linear regression model of the form

$$
\begin{aligned}
y_i &= w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + w_4 x_{i4} + \epsilon_i && (4)\\
x_{i1} &= 1 && (5)\\
x_{i2} &= g_i && (6)\\
x_{i3} &= \text{age of person i} && (7)\\
x_{i4} &= x_{i2} \times x_{i3} && (8)
\end{aligned}
$$

where $g_i = 1$ if person i is in some control group, and $g_i = 0$ otherwise. So we have

$$
\begin{aligned}
\mathbb{E}\left[y|\mathbf{x}_i, g_i = 0\right] &= w_1 + w_3 \text{age}_i && (9)\\
\mathbb{E}\left[y|\mathbf{x}_i, g_i = 1\right] &= (w_1 + w_2) + (w_3 + w_4)\text{age}_i && (10)
\end{aligned}
$$

Hence the difference in offsets between the two groups is w_2, and the difference in slopes is w_4. Sketch the regression line for the two groups assuming $w_1 = 1$, $w_3 = 1$ and with the following settings for the other parameters: (1) $w_2 = 0$, $w_4 = 0$, (2) $w_2 = 0$, $w_4 = 1$, (3) $w_2 = 1$, $w_4 = 0$, (4) $w_3 = 1$, $w_4 = 1$. You should have 4 figures, each with 2 lines. You can draw the figures by hand, or use Matlab. Assume the age ranges from 0 to 10.

3 Variance of a sum

Show that the variance of a sum is

$$\mathrm{var}\left[X+Y\right] = \mathrm{var}\left[X\right] + \mathrm{var}\left[Y\right] + 2\mathrm{cov}\left[X,Y\right] \tag{11}$$

where $\mathrm{cov}\left[X,Y\right]$ is the covariance between X and Y

4 Correlation coefficient is between -1 and +1

Prove that $-1 \leq \rho(X,Y) \leq 1$

5 Correlation coefficient for linearly related variables is ± 1

Show that, if $Y = aX + b$ for some parameters $a > 0$ and b, then $\rho(X, Y) = 1$. Similarly show that if $a < 0$, then $\rho(X, Y) = -1$.

6 Uncorrelated does not imply independent

Let $X \sim U(-1, 1)$ and $Y = X^2$. Clearly Y is dependent on X (in fact, Y is uniquely determined by X). However, show that $\rho(X, Y) = 0$. Hint: if $X \sim U(a, b)$ then $E[X] = (b - a)/2$ and $\text{var}\,[X] = (b - a)^2/12$.

7 Bayes rule for medical diagnosis

After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of tetsing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. What are the chances that you actually have the disease? (Show your calculations as well as giving the final result.)

8 Conditional independence

1. Let $H \in \{1, \ldots, K\}$ be a discrete random variable, and let e_1 and e_2 be the observed values of two other random variables E_1 and E_2. Suppose we wish to calculate the vector

   $$\vec{P}(H|e_1, e_2) = (P(H = 1|e_1, e_2), \ldots, P(H = K|e_1, e_2))$$

   Which of the following sets of numbers are sufficient for the calculation?

   (a) $P(e_1, e_2)$, $P(H)$, $P(e_1|H)$, $P(e_2|H)$

   (b) $P(e_1, e_2)$, $P(H)$, $P(e_1, e_2|H)$

   (c) $P(e_1|H)$, $P(e_2|H)$, $P(H)$

2. Now suppose we now assume $E_1 \perp E_2 | H$ (i.e., E_1 and E_2 are conditionally independent given H). Which of the above 3 sets are sufficent now?

Show your calculations as well as giving the final result. Hint: use Bayes rule.

9 The Monty Hall problem

On a game show, a contestant is told the rules as follows:

> There are three doors, labelled 1, 2, 3. A single prize has been hidden behind one of them. You get to select one door. Initially your chosen door will *not* be opened. Instead, the gameshow host will open one of the other two doors, and *he will do so in such a way as not to reveal the prize.* For example, if you first choose door 1, he will then open one of doors 2 and 3, and it is guaranteed that he will choose which one to open so that the prize will not be revealed.
>
> At this point, you will be given a fresh choice of door: you can either stick with your first choice, or you can switch to the other closed door. All the doors will then be opened and you will receive whatever is behind your final choice of door.

Imagine that the contestant chooses door 1 first; then the gameshow host opens door 3, revealing nothing behind the door, as promised. Should the contestant (a) stick with door 1, or (b) switch to door 2, or (c) does it make no difference? You may assume that initially, the prize is equally likely to be behind any of the 3 doors. Hint: use Bayes rule.

10 Moments of a Bernoulli distribution

Let $X \in \{0, 1\}$ be a binary random variable (e.g., a coin toss). Suppose $p(X = 1) = \theta$. Then

$$p(x|\theta) = \mathrm{Ber}(X|\theta) = \theta^x (1-\theta)^{1-x} \tag{12}$$

is called a Bernoulli distribution. Prove the following facts:

$$\mathbb{E}[X] = p(X = 1) = \theta, \quad \mathrm{var}[X] = \theta(1-\theta) \tag{13}$$

11 Exchangeability

Consider Polya's urn. This is an urn containing r red balls and b blue balls. Now consider the following experiment: we draw a ball, note its color, and replace the ball back in the urn along with c additional balls of the same color. Let us denote the event of observing a red ball at the i^{th} trial by R_i, and similarly observing a blue ball by B_i. Let $X_i \in \{R, B\}$ be the random variable representing the color of the i'th ball, where $p(X_i = R) = p(R_i)$ and $p(X_i = B) = p(B_i)$. Prove that the X_i are exchangeable but not iid. (Hint: the X_i are identically distributed, but are not independent.)

12 Pairwise independence does not imply mutual independence

We say that two random variables are pairwise independent if

$$p(X_2|X_1) = p(X_2) \tag{14}$$

and hence

$$p(X_2, X_1) = p(X_1)p(X_2|X_1) = p(X_1)p(X_2) \tag{15}$$

We say that n random variables are mutually independent if

$$p(X_i|X_S) = p(X_i) \quad \forall \subseteq \{1, \ldots, n\} \setminus \{i\} \tag{16}$$

and hence

$$p(X_{1:n}) = \prod_{i=1}^{n} p(X_i) \tag{17}$$

Show that pairwise independence between all pairs of variables does not necessarily imply mutual independence. It suffices to give a counter example.

13 Conditional independence iff joint factorizes

In the text we said $X \perp Y|Z$ iff

$$p(x, y|z) = p(x|z)p(y|z) \tag{18}$$

for all x, y, z such that $p(z) > 0$. Now prove the following alternative definition: $X \perp Y|Z$ iff there exist function g and h such that

$$p(x, y|z) = g(x, z)h(y, z) \tag{19}$$

for all x, y, z such that $p(z) > 0$.

14 Deriving the inverse gamma density

Let $X \sim \mathrm{Ga}(a, b)$ and $Y = 1/X$. Show that $Y \sim \mathrm{IG}(a, b)$. Hint: use the change of variables formula.

15 Marginalizing a Dirichlet

Suppose

$$\begin{aligned} p(\theta_1, \theta_2, \theta_3) &= \mathrm{Dir}(\alpha_1, \alpha_2, \alpha_3) && (20) \\ &\propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} (1 - \theta_1 - \theta_2)^{\alpha_3 - 1} && (21) \end{aligned}$$

since $\theta_1 + \theta_2 + \theta_3 = 1$. Derive an expression for

$$p(\theta_1) = \int_0^{1-\theta_1} p(\theta_1, \theta_2) d\theta_2 \tag{22}$$

You may ignore (cancel) any normalizing constants in your final answer. But you should identify the functional form of the marginal, along with its parameters.

Hint 1: make the substitution $u = \frac{\theta_2}{1-\theta_1}$ so $1 - u = \frac{1-\theta_1-\theta_2}{1-\theta_1}$. The answer should be another Dirichlet. (In fact, it will be a beta distribution, which is just a special case of Dirichlet.)

Hint 2: Use the fact that

$$\int_0^1 x^{a-1}(1-x)^{b-1} dx = B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{23}$$

since this is the normalization constant of the beta distribution.

Hint 3: Using the above substitutions, rewrite the integral as

$$p(\theta_1) \propto \theta_1^a (1 - \theta_1)^b \int_0^1 u^c (1-u)^d du \tag{24}$$

for suitably chosen a, b, c, d. (You must derive this equation and work out what these values are!)

16 Student T as infinite mixture of Gaussians

It turns out that several distributions of interest can be expressed as an "infinite" weighted sum of Gaussians, where each Gaussian has a different variance, i.e.,

$$p(x) = \int \mathcal{N}(x|\mu, \tau^2)\pi(\tau^2)d\tau^2 \tag{25}$$

for some distribution $\pi(\tau^2)$. This is called a **Gaussian scale mixture**.

Show that the Student distribution can be written as follows:

$$\mathcal{T}(x|\mu, \sigma^2, \nu) = \int_0^\infty \mathcal{N}(x|\mu, \sigma^2/\lambda)\mathrm{Ga}(\lambda|\frac{\nu}{2}, \frac{\nu}{2})d\lambda \tag{26}$$

17 Normalization constant for a 1D Gaussian

The normalization constant for a zero-mean Gaussian is given by

$$Z = \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \tag{27}$$

where $a = -\infty$ and $b = \infty$. To compute this, consider its square

$$Z^2 = \int_a^b \int_a^b \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) dxdy \tag{28}$$

Let us change variables from cartesian (x, y) to polar (r, θ) using $x = r\cos\theta$ and $y = r\sin\theta$. Since $dxdy = rdrd\theta$, and $cos^2\theta + \sin^2\theta = 1$, we have

$$Z^2 = \int_0^{2\pi} \int_0^{\infty} r\exp\left(-\frac{r^2}{2\sigma^2}\right) drd\theta \tag{29}$$

Evaluate this integral and hence show $Z = \sigma\sqrt{(2\pi)}$. Hint 1: separate the integral into a product of two terms, the first of which (involving $d\theta$) is constant, so is easy. Hint 2: if $u = e^{-r^2/2\sigma^2}$ then $du/dr = -\frac{1}{\sigma^2} r e^{-r^2/2\sigma^2}$, so the second integral is also easy (since $\int u'(r)dr = u(r)$).

18 Uncorrelated does not imply independent unless *jointly* Gaussian

Let $X \sim \mathcal{N}(0,1)$ and $Y = WX$, where $p(W = -1) = p(W = 1) = 0.5$. It is clear that X and Y are not independent, since Y is a function of X.

1. Show $Y \sim \mathcal{N}(0,1)$. Thus X and Y are both Gaussian. Hint: To show the mean is zero, use the fact that X and W are independent. To show the variance is 1, use the **rule of iterated variance**

$$\text{var}[Y] = \mathbb{E}[\text{var}[Y|W]] + \text{var}[\mathbb{E}[Y|W]] \tag{30}$$

2. Show $\text{cov}[X,Y] = 0$. Thus X and Y are uncorrelated but dependent, even though they are Gaussian. Hint: use the definition of covariance

$$\text{cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{31}$$

and the **rule of iterated expectation**

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|W]] \tag{32}$$