10-701/15-781, Machine Learning: Homework 1

Eric Xing, Tom Mitchell, Aarti Singh
Carnegie Mellon University
Updated on January 12, 2010

1 Probabilities [30 pt, Field Cady]

Probability is, in many ways, the most fundamental mathematical technique for machine learning. This problem will review several basic notions from probability and make sure that you remember how to do some elementary proofs.

Recall that for a discrete random variable (r.v.) X whose values are integers, we frequently use the notation $P(X = x)$ for the probability its value is x. If the r.v. Y is *continuous*, we typically use a "density function" $p(Y = y)$. The conditions for $P(X = x)$ to be a valid probability distribution are that $\sum_{-\infty}^{\infty} p(X = x) = 1$ and $P(X = x) \geq 0 \forall x$. Similarly for $p(Y = y)$ to be a valid continuous distribution, $\int_{-\infty}^{\infty} p(Y = y)dy = 1$ and $p(Y = y) \geq 0$.

Sometimes the underlying probability space has more than one variable (for example, the height and weight of a person). In this case, we may use notation like $p(X = x, Y = y)$ to denote the probability density function in several dimensions.

1.1 Expectations [10 pt]

Expectation is another word for "mean". They are similar to "average"; the difference is that an average usually refers to the average of some data we collected, whereas expectation usually refers to the underlying distribution from which we have sampled. For a discrete r.v. X, the expectation value is defined to be $E[X] = \sum_{-\infty}^{\infty} iP(X = i)$. If Y is a continuous random variable, $E[Y] = \int_{-\infty}^{\infty} yp(y)dy$.

1. Show that, for discrete r.v. W and Z, $E[W + Z] = E[W] + E[Z]$.

2. Show that, for *continuous* r.v. W and Z, $E[W + Z] = E[W] + E[Z]$.

1.2 Solution

$$\begin{aligned}
&\text{Discrete} \\
E[W+Z] &= \Sigma_{eventse}(W+Z)P(e) \\
&= \Sigma_{eventse}WP(e) + \Sigma_{eventse}ZP(e) \\
&= E[W] + E[Z] \\
&\text{continuous} \\
E[W+Z] &= \int_w \int_z (w+z) f_{WZ}(w,z)dzdw \\
&= \int_w \int_z w f_{WZ}(w,z)dzdw + \int_w \int_z z f_{WZ}(w,z)dzdw \\
&= \int_w w \int_w f_{WZ}(w,z)dzdw + \int_z z \int_w f_{WZ}(w,z)dzdw \\
&= \int_w w f_W(w)dw + \int_z f_Z(z)dz \\
&= E[W] + E[Z]
\end{aligned}$$

1.3 Independence [10 pt]

Intuitively, two r.v. X and Y are "independent" if knowledge of the value of one tells you nothing at all about the value of the other. Precisely, if X and Y are discrete, independence means that $P(X=x, Y=y) = P(X=x)P(Y=y)$, and if they are continuous, $p(X=x, Y=y) = p(X=x)p(Y=y)$. Show the following, for *independent* r.v. X and Y:

1. If X and Y are discrete, $E[XY] = E[X]E[Y]$.
2. If X and Y are *continuous*, $E[XY] = E[X]E[Y]$.

1.4 Solution

$$\begin{aligned}
&\text{Discrete} \\
E[XY] &= \Sigma_{X=x}\Sigma_{Y=y} xyP(X = x \& Y = y) \\
&= \Sigma_{X=x}\Sigma Y = yxyP(X = x)P(Y = y) \\
&= \Sigma_{X=x} xP(X = x)\left(\Sigma Y = yyP(Y = y)\right) \\
&= \Sigma_{X=x} xP(X = x)E[Y] \\
&= E[Y]\Sigma_{X=x} xP(X = x) \\
&= E[Y]E[X] \\
&\text{continuous} \\
E[XY] &= \int_x \int yxyf_{XY}(x, y)dydx \\
&= \int_x \int yxyf_X(x)f_Y(y)dydx \\
&= \int_x xf_X(x)\left(\int yyf_Y(y)dy\right)dx \\
&= \int_x xf_X(x)\left(E[Y]\right)dx \\
&= E[Y]\int xf_X(x)dx \\
&= E[Y]E[X]
\end{aligned}$$

1.5 Variance [10 pt]

Variance for a r.v. X indicates how "spread out" the distribution is. Precisely, if $\bar{X} = E[X]$, the variance is defined to be $Var[X] = E[(X - \bar{X})^2]$. Show the following:

1. For a (discrete *or* continuous) random variable X, $Var[X] = E[X^2] - (E[X])^2$.

   Hint: you don't have to treat the discrete and continuous cases separately; it can be done just using expectation.

2. Let X be continuous, and let it follow the celebrated normal distribution: $p(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x-\mu)^2/(2\sigma^2)}$, where σ can be any positive real number, and μ can be any real number. Show that $Var[X] = \sigma^2$.

1.6 Solution

$$\begin{aligned}
Var[X] &= E[(X - E[x])^2] \\
&= E[X^2 - 2XE[X] + E[X]^2] \\
&= E[X^2] + E[-2E[X]X] + E[X]^2 \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - (E[X])^2
\end{aligned}$$

For the normal distribution, we see the distribution is symmetric about $x = \mu$. To calculate variance, without loss of generality we can assume that $\mu = 0$. Then

$$\begin{aligned}
Var[X] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 e^{-x^2/(2\sigma^2)} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int (\sigma u)^2 e^{-u^2/2} (\sigma du) \\
&\quad \text{change of vars } u = x/\sigma \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int u^2 e^{-u^2/2} du \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[\int (-u)(-u e^{-u^2/2}) du \right] \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[(-u)(e^{-u^2/2})|_{-\infty}\infty - \int (e^{-u^2/2})(-1) du \right] \\
&\quad \text{integration by parts} \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[0 + \int e^{-u^2/2} du \right] \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} \\
&= \sigma^2
\end{aligned}$$

where we used the following very clever trick:

$$\begin{aligned}
\left(\int_{u=-\infty} \infty e^{-u^2/2} du \right)^2 &= \left(\int e^{-x^2/2} dx \right) \left(\int e^{-y^2/2} dy \right) \\
&= \int_{x=-\infty}^{\infty} \int y = -\infty^{\infty} e^{-(x^2+y^2)/2} dy dx \\
&= \int \int_{plane} e^{(x^2+y^2)/2} dy dx \\
&= \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} e^{-r^2/2} (r dr d\theta) \\
&\quad \text{changing to polar coordinates} \\
&= 2\pi \int_{r=0} \infty e^{-r^2/2} r dr \\
&= 2\pi \int_{v=0}^{\infty} e^{-v} dv \text{setting } v = r^2/2 \\
&= 2\pi \left[-e^{-v} \right]_0^{\infty} \\
&= 2\pi(0 - (-1)) \\
&= = 2\pi
\end{aligned}$$