CS-XXX: Introduction to Machine Learning

Statistics

Arbish Akram

Department of Computer Science
Government College University

Slides by Dr. Nazar Khan

Expectation

- Expectation is a weighted average of a function.
- Weights are given by $p(x)$.

$$\mathbb{E}[f] = \sum_x p(x) f(x) \qquad \longleftarrow \text{For discrete } x$$

$$\mathbb{E}[f] = \int_x p(x) f(x) dx \qquad \longleftarrow \text{For continuous } x$$

- When data is finite, expectation $\approx$ ordinary average. Approximation becomes exact as $N \to \infty$ (*Law of large numbers*).

Expectation

- Expectation of a function of several variables

$$\mathbb{E}\left[f(x,y)\right] = \sum_{x,y} p(x,y) f(x,y)$$

- Expectation with respect to one variable

$$\mathbb{E}_x\left[f(x,y)\right] = \sum_{x} p(x) f(x,y) \qquad \text{(function of } y\text{)}$$

- *Conditional expectation*

$$\mathbb{E}_{x|y}\left[f\right] = \sum_{x} p(x|y) f(x)$$

Variance

- *Variance* measures variability of a random variable around its mean.

$$
\begin{aligned}
var\left[f\right] &= \mathbb{E}\left[\left(f(x) - \mathbb{E}\left[f(x)\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(f(x)\right)^2\right] - \mathbb{E}\left[f(x^2)\right]
\end{aligned}
$$

- On average, how far does a random variable stay from its mean?

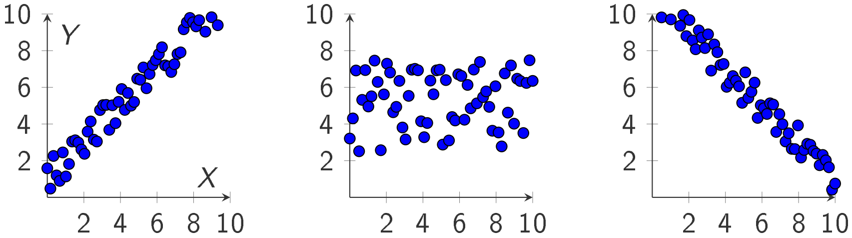
Covariance
Univariate

- For 2 univariate random variables occuring in pairs (x, y), **covariance** expresses how much x and y vary together.

$$
\begin{aligned}
cov[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\
&= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\,\mathbb{E}[y]
\end{aligned}
$$

- For independent random variables x and y, $cov[x, y] = 0$. Why?

Positive Covariance **Zero Covariance** **Negative Covariance**



Covariance
Multivariate

- For multivariate random variables $\mathbf{x} \in \mathbb{R}^D$ and $\mathbf{y} \in \mathbb{R}^K$, $cov[\mathbf{x}, \mathbf{y}]$ is a $D \times K$ matrix.
- Expresses how each element of $\mathbf{x}$ varies with each element of $\mathbf{y}$.

$$
\begin{aligned}
cov[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y} - \mathbb{E}[\mathbf{y}]\}^T\right] \\
&= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\mathbf{x}\mathbf{y}^T\right] - \mathbb{E}[\mathbf{x}]\,\mathbb{E}[\mathbf{y}]^T \\
&= \begin{bmatrix} cov[x_1, y_1] & cov[x_1, y_2] & \cdots & cov[x_1, y_K] \\ cov[x_2, y_1] & cov[x_2, y_2] & \cdots & cov[x_2, y_K] \\ \vdots & \vdots & \ddots & \vdots \\ cov[x_D, y_1] & cov[x_D, y_2] & \cdots & cov[x_D, y_K] \end{bmatrix}
\end{aligned}
\tag{1}
$$

Covariance
Multivariate

- Covariance of multivariate $\mathbf{x}$ with itself can be written as $cov\left[\mathbf{x}\right] \equiv cov\left[\mathbf{x}, \mathbf{x}\right]$.
- $cov\left[\mathbf{x}\right]$ expresses how each element of $\mathbf{x}$ varies with every other element.

$$cov[\mathbf{x}] = \begin{bmatrix} var[x_1] & cov[x_1, x_D] & \cdots & cov[x_1, x_D] \\ cov[x_2, x_1] & var[x_2] & \cdots & cov[x_2, x_D] \\ \vdots & \vdots & \ddots & \vdots \\ cov[x_D, x_1] & cov[x_D, x_2] & \cdots & var[x_D] \end{bmatrix} \tag{2}$$

Bayesian View of Probability

- So far we have considered probability as the *frequency of random, repeatable events*.
- What if the events are not repeatable?
  - Was the moon once a planet?
  - Did the dinosaurs become extinct because of a meteor?
  - Will the ice on the North Pole melt by the year 2100?
- For non-repeatable, yet uncertain events, we have the **Bayesian view** of probability.

Bayesian View of Probability

$$p(\mathbf{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})}$$

- Measures the uncertainty in model $\mathbf{w}$ after observing the data $\mathcal{D}$.
- This uncertainty is measured via conditional $p(\mathcal{D}|\mathbf{w})$ and prior $p(\mathbf{w})$.
- Treated as a function of $\mathbf{w}$, the conditional probability $p(\mathcal{D}|\mathbf{w})$ is also called the **likelihood function**.
- Expresses how likely the observed data is for any given model $\mathbf{w}$.