Train In Data

Joint Distribution and Conditional Probability

Data – Counts

	German Sheppard	Dobermann	Labrador	Marginal (Dysplasia)
None	10	30	60	100
Mild	30	40	210	280
Severe	170	30	30	230
Marginal (Breed)	200	100	300	600

- Dog breeds as columns
- The degree of hip dysplasia as rows

Data – Frequency, Probability

	German Sheppard	Dobermann	Labrador	Marginal (Dysplasia)
None	0.017	0.050	0.100	0.167
Mild	0.050	0.067	0.350	0.467
Severe	0.283	0.050	0.050	0.383
Marginal (Breed)	0.333	0.167	0.500	1.000

- Dog breeds as columns
- The degree of hip dysplasia as rows

Marginal Probability

	German Sheppard	Dobermann	Labrador	Marginal (Dysplasia)
None	0.017	0.050	0.100	0.167
Mild	0.050	0.067	0.350	0.467
Severe	0.283	0.050	0.050	0.383
Marginal (Breed)	0.333	0.167	0.500	1.000

- Probability that we select a dog of our population and it is a German Sheppard.
- P(breed) ➔ marginal probability

Marginal Probability

	German Sheppard	Dobermann	Labrador	Marginal (Dysplasia)
None	0.017	0.050	0.100	0.167
Mild	0.050	0.067	0.350	0.467
Severe	0.283	0.050	0.050	0.383
Marginal (Breed)	0.333	0.167	0.500	1.000

- Probability that we select a dog of our population and it shows mild dysplasia.
- P(dysplasia) ➔ marginal probability

Joint Probability

	German Sheppard	Dobermann	Labrador	Marginal (Dysplasia)
None	0.017	0.050	0.100	0.167
Mild	0.050	0.067	0.350	0.467
Severe	0.283	0.050	0.050	0.383
Marginal (Breed)	0.333	0.167	0.500	1.000

- Probability that we select a dog of our population and it is a Dobermann with mild dysplasia.
- P(breed, dysplasia) ➔ joint probability

Joint Probability

- The probability of particular combinations of 2 events taking place.
- The joint probability is symmetric ➔ P(A, B) = P(B, A)
- In our example, P(breed, dysplasia) = P(dysplasia, breed)

Marginal Probability

- The marginal probability is the probability of particular event, collapsed across the values of the other event.
- The marginal probability is the sum of the joint probabilities.
- $P(A) = \sum P(A, B)$
- In our example, $P(breed) = \sum P(dysplasia, breed)$

Conditional Probability

	German Sheppard
None	0.017
Mild	0.050
Severe	0.283
Marginal (Breed)	0.333

Divide by 0.333

	German Sheppard
None	0.050
Mild	0.150
Severe	0.851
Marginal (Breed)	1

- A German Sheppard comes to the vet.
- We wonder: what is the probability of dysplasia given that it is a German Sheppard?

Conditional Probability

- Probability of 1 outcome, given that the other is true.
- P(A | B) ➔ probability of A given B
- P(A | B) = P(A, B) / P(B)
- Conditional probability is not symmetric ➔ P(A | B) ≠ P(B | A)
- In our example, $P(\text{dysplasia} \mid \text{breed}) = \frac{P(\text{dysplasia, breed})}{P(\text{breed})}$

Prior and Posterior Probability

- Prior probability is the unconditional probability assigned to an event before any relevant information is taken into account.
  - In our example, P(dysplasia)
- The posterior probability of an event, is the **conditional probability** that is assigned after taking into account the new evidence.
  - In our example, P(dysplasia | breed = German Sheppard)
- Prior and posterior are mathematically related by Bayes' Rule

Train In Data

THANK YOU

www.trainindata.com