

Model Contingency [2] of World on Data (Generative)

Qi Zhao

July 24, 2018

1. Introduction

The two model types result in different approaches to inference, which can be used to estimate the world state w from an observed data example x , based on modeling the posterior $\Pr(w|x)$ and the likelihood $\Pr(x|w)$, respectively. The models were carefully chosen so that they predict exactly the same posterior $P(w|x)$ over the world state. For the generative model [3], it computes the posterior using Bayes' rule [6], which sometimes results in complex inference algorithms.

2. Descriptions

In the generative formulation, it need choose a probability distribution [5] over the data x and make its parameters contingent on the world state w . Since the data are univariate and continuous, it will model the data as a normal distribution with fixed variance, σ^2 and a mean μ that is a linear function $\phi_0 + \phi_1 w$ of the world state (Figure 1) in Equation 1.

$$\Pr(x|w, \theta) = \text{Norm}_x[\phi_0 + \phi_1 w, \sigma^2] \quad (1)$$

It will choose a univariate normal and allow the variance σ^2 and the mean μ to be functions of the binary world state w (Figure 2). So the likelihood is in Equation 2.

$$\Pr(x|w, \theta) = \text{Norm}_x[\mu_w, \sigma_w^2] \quad (2)$$

3. Conclusions

For binary classification [4], there is an asymmetry between the world state, which is discrete, and the measurements, which are continuous. Consequently, the generative and discriminative models look quite different, and the posteriors over the world state w as a function of the data x have different shapes. For the discriminative model [1], this function is by definition sigmoidal, but for the generative case it has a more complex form that was implicitly defined by the normal likelihoods. In general, choosing to model $\Pr(w|x)$ or $\Pr(x|w)$ will affect the expressiveness of the final model.

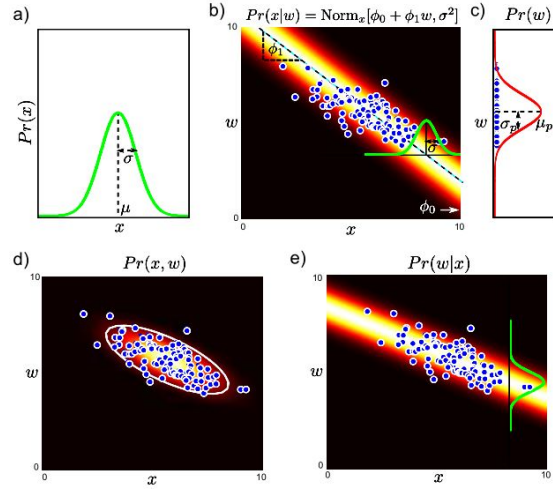


Figure 1. Regression by modeling likelihood $\Pr(x|w)$ (generative). a) We represent the data x with a normal distribution. b) We make the normal parameters functions of the world state w . Here the mean is a linear function $\mu = \phi_0 + \phi_1 w$ of the world state and the variance σ^2 is fixed. The learning algorithm fits the parameters $\theta = \{\phi_0, \phi_1, \sigma^2\}$ to example training pairs $x_i, w_{i=1}^T$ (blue dots). c) We also learn a prior distribution over the world state w (here modeled as a normal distribution with parameters $\theta_p = \{\mu_p, \sigma_p\}$). In inference we take a new datum x and compute the posterior $\Pr(w|x)$ over the state. d) This can be done by computing the joint distribution $\Pr(x, w) = \Pr(x|w)\Pr(w)$ (weighting each row of (b) by the appropriate value from the prior) and e) normalizing the columns $\Pr(w|x) = \Pr(x, w)/\Pr(x)$. Together these operations implement Bayes rule: $\Pr(w|x) = \Pr(x|w)\Pr(w)/\Pr(x)$.

References

- [1] C. Desai, D. Ramanan, and C. C. Fowlkes. Discriminative models for multi-class object layout. In *IEEE International Conference on Computer Vision*, pages 229–236, 2010. 1
- [2] F. E. Fiedler. A contingency model of leadership effectiveness. *Advances in Experimental Social Psychology*, 1(1):149–190, 1964. 1
- [3] T. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Advances in Neural Information Processing Systems*, pages 487–493, 1999. 1
- [4] L. Li and H.-T. Lin. Ordinal regression by extended binary

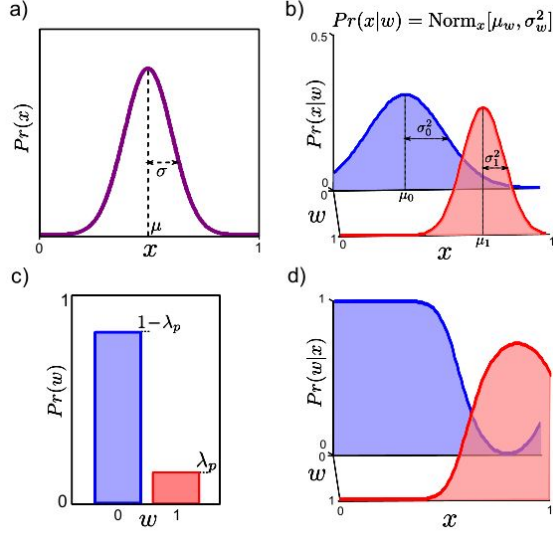


Figure 2. Classification by modeling the likelihood $Pr(x|w)$ (generative). a) We choose a normal distribution to represent the data x . b) We make the parameters μ , σ^2 of this normal a function of the world state w . In practice, this means using one set of mean and variance parameters when the world state $w = 0$ and another when $w = 1$. The learning algorithm fits the parameters $\theta = \{\mu_0, \mu_1, \sigma_0^2, \sigma_1^2\}$ to example training pairs $\{x_i, w_i\}_{i=1}^I$. c) We also model the prior probability of the world state w with a Bernoulli distribution with parameter λ_p . d) In inference we take a new datum x and compute the posterior $Pr(w|x)$ over the state using Bayes rule.

classification. In *Advances in Neural Information Processing Systems*, pages 865–872, 2007. [1](#)

- [5] B. Rannala and Z. Yang. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution*, 43(3):304–311, 1996. [1](#)
- [6] R. A. Waller and D. B. Duncan. A Bayes’ rule for the symmetric multiple comparisons problem. *Journal of the American Statistical Association*, 64(328):1484–1503, 1969. [1](#)