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

Qi Zhao

July 22, 2018

1. Introduction

At an abstract level, the goal of computer vision problems is to use the observed image data to infer something about the world. For example, it might observe adjacent frames [5] of a video sequence and infer the camera motion, or it might observe a facial image and infer the identity. There are two distinct approaches to modeling the relationship between the world state w and the data x, corresponding to modeling the posterior Pr(w|x) [3], or the likelihood Pr(x|w) [1]. The two model types result in different approaches to inference. For the discriminative model [4], it describe the posterior Pr(w|x) directly.

2. Descriptions

It defines a probability distribution over the world state w and make its parameters contingent on the data x. Since the world state is univariate and continuous, it chose the univariate normal. It fixes the variance, σ^2 and make the mean μ a linear function ϕ_0 + ϕ_1x of the data. So it can be seen in Equation 1.

$$Pr(w|x,\theta) = Norm_x[\phi_0 + \phi_1 x, \sigma^2] \tag{1}$$

where θ = ϕ_0, ϕ_1, σ^2 are the unknown parameters of the model (Figure 1). This model is referred to as linear regression.

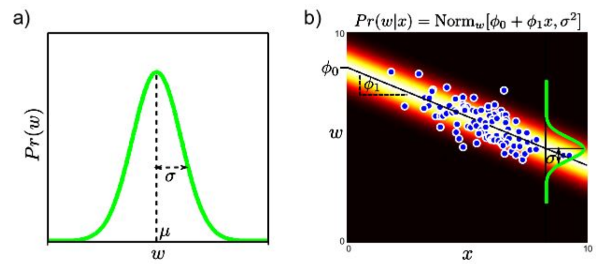

Figure 1. Regression by modeling the posterior Pr(w|x) (discriminative). a) We model the world state w as normally distributed. b) We make the normal parameters a function of the observations x: the mean is a linear function μ = ϕ_0 + ϕ_1x of the observations, and the variance 2 is fixed. The learning algorithm fits the parameters θ = ϕ_0, ϕ_1, σ^2 to example training pairs x_i, $w_i{}_{i=1}^{I}$ (blue dots). In inference we take a new observation x and compute the posterior distribution Pr(w|x) over the state.

The learning algorithm estimates the model parameters θ from paired training examples x_i, $w_i{}_{i=1}^{I}$. For example, in the MAP approach [6], it need seek in Equation 2.

$$\begin{aligned}\hat{\theta} &= argmax_\theta[Pr(\theta|w_{1\ldots I}, x_{1\ldots I})] \\ &= argmax_\theta[Pr(w_{1\ldots I}|x_{1\ldots I}, \theta)Pr(\theta)] \\ &= argmax_\theta[\prod_{i=1}^{I} Pr(w_i|x_i, \theta)Pr(\theta)]\end{aligned} \tag{2}$$

where assumed that the I training pairs x_i, $w_i{}_{i=1}^{I}$ are independent, and defined a suitable prior Pr(θ).

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977. 1

[2] F. E. Fiedler. A contingency model of leadership effectiveness. *Advances in Experimental Social Psychology*, 1(1):149–190, 1964. 1

[3] J. Hinchey, C. Chaves, B. Appignani, J. Breen, L. Pao, A. Wang, M. S. Pessin, C. Lamy, J. L. Mas, and L. R. Caplan. A reversible posterior leukoencephalopathy syndrome. *New England Journal of Medicine*, 334(8):494–500, 1996. 1

[4] T. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Advances in Neural Information Processing Systems*, pages 487–493, 1999. 1

[5] J. Y. Lin, C. W. Lee, and Z. Chen. Identification of business forms using relationships between adjacent frames. *Machine Vision & Applications*, 9(2):56–64, 1996. 1

[6] H. Shen, L. Zhang, B. Huang, and P. Li. A MAP approach for joint motion estimation, segmentation, and super resolution. *IEEE Transactions on Image Processing*, 16(2):479–490, 2007. 1