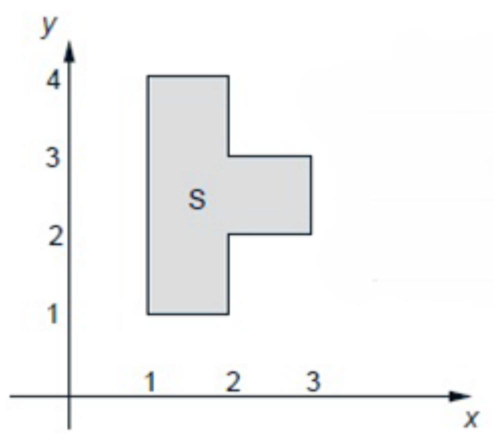
1. Derive the maximum likelihood estimate of the noise variance σ^2 in the linear regression model. (Hint: Formulate loglikelihood in terms of σ^2 and also take derivative wrt σ^2, NOT σ).

2. Imagine binary Naïve Bayes model where the input X is d dimensional and each of the dimensions of X is categorical taking one of K values (instead of being binary as in the lecture). Let η_{jlk} represent the probability that dimension j of X takes value k under class l, where l is either 0 or 1. Derive the MAP estimate for η_{jlk}. (Hint: Use Dirichlet as prior for Multinomial.)

3. Derive the form of the gradient of the likelihood of a binary logistic regression model. (Hint: Write $\mu_n = \frac{1}{1+e^{-\eta_n}}$ and $\eta_n = w^T x_n$. Then use $\frac{dl(w)}{dw} = \frac{dl(w)}{d\mu}\frac{d\mu}{d\eta}\frac{d\eta}{dw}$)

4. If Θ be a parameter and $\widehat{\Theta}$ be its estimate, then show that MSE($\widehat{\Theta}$) = Var($\widehat{\Theta}$) + Bias($\widehat{\Theta}$)^2

5. Find conditional distributions $f_{X|Y}()$ for the following joint distribution



6. Show that the geometric distribution is memoryless. Specifically, if X be a geometrically distributed random variable, then show that $P(X \geq t + s | X \geq t) = P(X \geq s)$

7. Derive the variance of a Binomial Distribution using the relationship between the Binomial and Bernoulli distributions.